UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

Dinâmica Molecular de Proteínas: estabilidade e renaturação
térmica

RICARDO OLIVEIRA DOS SANTOS SOARES

Ribeirão Preto
2009

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

Dinâmica Molecular de Proteínas: estabilidade e renaturação
térmica

Dissertação de Mestrado apresentado ao
Programa de Pós-Graduação em Ciências
Farmacêuticas, para a obtenção do título de
Mestre em Ciências Farmacêuticas.

Área de concentração: Física Biológica

Orientado: Ricardo O. S. Soares
Orientador: Prof. Dr. Antonio Caliri

Ribeirão Preto
2009

SOARES, R.O.S. **Dinâmica Molecular de Proteínas: estabilidade e renaturação térmica**. 2009. 86 p. Dissertação (Mestrado) – Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

Errata

Folha	Linha	Onde se lê	Leia-se
1	1 e 2	peso atômico	massa molecular
2	6	precisa	-
2	10	conformada espacialmente	enovelada
4	5	predição	aplicação
4	24	peptídeos	nucleotídeos
13	Figura 1.3	[33]	(Wallner <i>et al.</i> , 2005)
13	Figura 1.3	[37]	(Sali & Blundell, 1993)
13	Figura 1.3	[43]	(Levitt, 1992)
13	Figura 1.3	[34]	(Petrey <i>et al.</i> , 2003)
16	7	4-a	1.4 (a)
18	3	abientes	ambientes
19	7	é	-
21	14	código genético	sequência genômica
23	19	deBrogie	deBroglie
46	9	(ref-1)	-
56	1	Coisa	Fato
62	2	[REF Lesk]	(Lesk, 2003)
62	16	se	ser
63	3	laterais	laterais
63	5	<i>re-foldind</i>	<i>re-folding</i>
63	16	encontramsua	encontram sua
63	20	hamvientais	ambientais

Adição das seguintes referências:

- Levitt, M. (1992). "Accurate Modeling of Protein Conformation by Automatic Segment Matching." Journal of Molecular Biology **226**(2): 507-533.
- Petrey, D., Z. X. Xiang, et al. (2003). "Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling." Proteins-Structure Function and Genetics **53**(6): 430-435.
- Sali, A. and T. L. Blundell (1993). "Comparative Protein Modeling by Satisfaction of Spatial Restraints." Journal of Molecular Biology **234**(3): 779-815.
- Wallner, B. and A. Elofsson (2005). "All are not equal: A benchmark of different homology modeling programs." Protein Science **14**(5): 1315-1327.

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Soares, Ricardo Oliveira dos Santos

Dinâmica Molecular de Proteínas: estabilidade e renaturação térmica/ Ricardo Oliveira dos Santos Soares; orientador Antonio Caliri – Ribeirão Preto, 2008.

86 p.; il.; 30cm

Dissertação (Mestrado – Programa de Pós-Graduação em Ciências Farmacêuticas. Área de Concentração: Física Biológica) – Faculdade de Ciências Farmacêuticas de Ribeirão Preto.

1. Estrutura de Proteínas. 2. *folding* de Proteínas. 3. Dinâmica Molecular. 4. Pontes dissulfeto. 5. Ligações de Hidrogênio. 6. Ts Kappa. 7. *Tityus serrulatus*.

Dinâmica Molecular de Proteínas: estabilidade e renaturação térmica

Dissertação de Mestrado apresentado ao
Programa de Pós-Graduação em Ciências
Farmacêuticas, para a obtenção do título de
Mestre em Ciências Farmacêuticas

Área de concentração: Física Biológica

Aprovado em: ____ / ____ / ____

Banca Examinadora

Prof(a). Dr(a).: _____

Instituição: _____

Assinatura: _____

Prof(a). Dr(a).: _____

Instituição: _____

Assinatura: _____

Prof(a). Dr(a).: _____

Instituição: _____

Assinatura: _____

Candidato: _____

Froh, wie seine Sonnen fliegen
Feliz, como vosso sol que vôa,
Durch des Himmels prächt'gen Plan,
Através do esplêndido Paraíso,
Wandelt, Brüder, eure Bahn,
Vague, irmãos, em seu caminho,
Freudig, wie ein Held zum Siegen.
Alegres, como um herói indo para a Vitória.

(Letra de Ode “An die Freude”, por Friederich von Schiller, 1759-1805)
(Versos cantados no quarto movimento da nona sinfonia de Beethoven)

DEDICATÓRIA

Gostaria de tomar este espaço para dedicar este trabalho à minha família e à ciência.

Ricardo O. S. Soares

AGRADECIMENTOS

Ao **Prof. Dr. Antonio Caliri**, pela orientação deste trabalho. Por ser mais que um orientador, um amigo. Pela luz que tem me trazido em meu aprendizado nestes últimos anos, não apenas acadêmico, mas também de vida.

À Universidade de São Paulo e à Faculdade de Ciências Farmacêuticas de Ribeirão Preto, pela oportunidade oferecida, pelos recursos técnicos e humanos que permitiram meu aprimoramento e a execução desse trabalho.

Ao Programa de Pós-Graduação desta Faculdade e à CAPES, pelo apoio financeiro nesse período, o qual permitiu que eu me concentrasse em minha pesquisa.

Aos funcionários da seção de Pós-Graduação desta Faculdade: Ana, Rosana, Rossana e Eleni, pela grande eficiência e pelo tratamento humano e gentil para comigo.

Aos professores das disciplinas pelas quais eu passei durante este mestrado, pelos ensinamentos.

Ao Prof. Dr. Fernando Luís Barroso da Silva, pela ajuda no contato bem sucedido com o Prof. Dr. David van der Spoel, o qual pôde vir ao Brasil para colaborar com um simpósio internacional que eu participei como comissão organizadora (III SINPOSPq).

Aos meus colegas do III SINPOSPq, que me proporcionaram um aprendizado extra-acadêmico sobre como funciona uma organização/administração de um grande evento científico.

Aos meus colegas de trabalho, agora amigos, pela companhia no dia a dia e em diversas situações agradáveis durante a convivência.

Aos meus grandes amigos de Piracicaba e de Ribeirão Preto, simplesmente por tudo que uma amizade traz, e pela ajuda em momentos difíceis. Agradeço especialmente Fernando Rossi Trigo, pelo apoio, companheirismo e amizade na vivência em nossa moradia estudantil desde a época da graduação.

À Márcia Cristina Vitti, uma pessoa muito especial para mim, maravilhosa em suas qualidades e defeitos.

À ciência, que me fascina, e à música, que me inspira.

Finalmente aos meus preciosíssimos pais, Wanderley e Isabel, eu não seria metade do que sou sem eles. Aos meus irmãos Marcelo e Cláudia, por existirem. Aos meus cunhados, agora irmãos, e aos meus queridos sobrinhos que fazem me sentir uma pessoa responsável e mais feliz.

Ricardo O. S. Soares

RESUMO

SOARES, R.O.S. Dinâmica Molecular de Proteínas: estabilidade e renaturação térmica. 2009. 86 p. Dissertação (Mestrado) – Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

Proteínas são heteropolímeros lineares essenciais à vida, responsáveis pela estruturação dos organismos e pela maioria dos processos bioquímicos que os mantêm vivos e permitem sua reprodução. Essa variedade de funções é refletida na diversidade estrutural encontrada no universo das proteínas, já que sua função é intrinsecamente ligada à sua rigorosa conformação espacial. A partir dos experimentos de Anfinsen (1973), ficou demonstrado que o enovelamento dessas moléculas (*folding*) se dá essencialmente por meio de um processo físico-químico guiado pela interação entre os aminoácidos da cadeia protéica e entre estes e o meio solvente, quando sob condições fisiológicas (temperatura, pressão, pH). O completo entendimento do mecanismo de *folding* tem também importância médica, pois várias doenças como mal de Alzheimer, diabetes tipo II, encefalite bovina espongiforme e várias formas de câncer estão relacionadas com falhas estruturais das proteínas.

Neste trabalho, por meio de experimentação computacional por dinâmica molecular (DM) em diferentes condições térmicas, estudamos inicialmente o papel das pontes dissulfeto (S-S) e das ligações de hidrogênio (LH) na estabilidade da proteína. Em seguida, adotando exclusivamente o regime de alta temperatura ($T = 448\text{K}$) em combinação com simulações de longa duração (até $\sim 100\text{ns}$), no intuito de expandir a exploração do espaço configuracional, verificamos a premissa de que as forças entrópicas, geradas pelo efeito hidrofóbico, seriam dominantes no processo de busca pela estrutura nativa. Neste trabalho foi utilizada como um protótipo de proteína pequena e com pontes S-S, a toxina Ts Kappa (MM=3,8 Kda; pdb id: 1tsk), que é dotada de três pontes S-S.

A estabilidade conformacional foi analisada por meio de uma série de simulações de DM em temperaturas crescentes e em duas situações: com e sem os *cross-links* S-S.

Nossos resultados indicam que para incrementos nas temperaturas significativamente elevadas, como 50K acima da temperatura em que a estrutura nativa foi determinada por NMR (283K), a remoção das S-S não compromete a estabilidade conformacional da proteína. De fato, a ausência dos *cross-links* elimina certas restrições geométricas permitindo agora que diferentes combinações de LH sejam feitas, inclusive entre resíduos adjacentes à cisteína, os quais de certa forma substituem as pontes S-S em seus papéis conformacionais pois a estrutura nativa é essencialmente mantida.

No segundo experimento o espaço configuracional foi varrido extensamente durante 100ns e à temperatura de 398K. No caso da Ts Kappa com suas pontes dissulfeto intactas, a desestruturação da proteína é limitada pelas fortes pontes covalentes S-S, mas com a remoção delas, a proteína se desnaturou completamente ao longo dos primeiros 50ns. Contudo, a partir deste ponto a cadeia desnaturada passou a seguir, de forma espontânea e sistemática, uma rota de re-estruturação em direção à nativa, com o re-estabelecimento de todas suas estruturas secundárias. Ao redor de 100ns a cadeia atingiu um estado de grande identidade estrutural com sua correspondente estrutura nativa.

Em conclusão, os presentes resultados corroboram as premissas de que o *folding* de proteínas ocorre por meio de um processo em duas etapas, temporalmente separadas: no início, as forças entrópicas são dominantes e são as que induzem a cadeia para a conformação nativa. Então, uma vez na vizinhança da estrutura nativa, as pontes de hidrogênio (agora protegidas da competição com o meio solvente), juntamente com um mais eficiente empacotamento estrutural das cadeias laterais devido às complementaridade estéricas das mesmas (e assim otimizando as interações de van der Waals), iniciam a etapa de estabilização energética da proteína.

Palavras-Chave: *folding* de proteína, estabilidade térmica, Ts Kappa, dinâmica molecular.

ABSTRACT

SOARES, R.O.S. **Protein Molecular Dynamics: stability and thermal renaturation.** (2009). 86 p. Dissertation (Master) – Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

Proteins are linear heteropolymers essential for life; they are responsible for many distinct functions as the structural components of organism, and for most of the biochemical processes to maintain a reproductive life. Such diversity of functions is correlated with the extremely large accessible conformational space, since function and spatial structure are interdependent. After Anfinsen experiments (1973), it becomes clear that the protein folding is essentially a physical-chemical process guided by interactions among the chain constituents (amino acid sequence) and interactions between the chain and the solvent, under physiological conditions (temperature, pressure, pH). Because miss-folded proteins are related with diseases (Alzheimer, type *II* diabetes, several forms of cancer, etc.) the full understanding of the folding mechanism has also significant medical interest.

In this work, by means of molecular dynamics (MD) simulations under distinct thermal conditions, we first consider the role of disulfide cross-links (S-S) and hydrogen bonds (HB) with respect to the protein thermal stability. Then, using exclusively high temperature regime ($T = 448\text{K}$) combined with extended time simulations (up to $\sim 100\text{ns}$), in order to fully span of configurational space, we analyzed the hypothesis that the entropic forces, generated by the hydrophobic effect, are dominant in the search process for the native structure. The protein Ts Kappa was used a prototype for small proteins having S-S bridges (MM=3,8Kda; 3 S-S - pdb id: 1tsk).

The thermal conformational stability was analyzed from a series of MD simulations under growing temperatures, using two distinct cases: with and without cross-links S-S. Our results suggest that for significant temperature increments, such as 50K above the temperature used in the Ts Kappa structure determination (by NMR at 283K), the thermal conformational stability of the proteins is not affected if the S-S bridges are removed. Indeed, cutting of the cross-links eliminates certain geometrical constraints, what permits the formation of new combinations of HB, which in some way take the place of the S-S bridges on its conformational role since the native structure is essentially maintained.

In the second computational experiment, the configurational space was extensively swapped during 100ns at a fixed temperature $T=398\text{K}$. In the case with preserved S-S bridges, the structural unpacking is limited by the three covalent cross-links, but without the S-S bridges the protein denaturation was complete after 50ns. However, after this point the chain started spontaneous and systematically a configurational rote that finally, after about 100ns, reached a conformation very similar with the native ($\text{RMSD} \approx 0.5\text{nm}$), reestablishing all its secondary structure.

Concluding, the present results corroborate the hypothesis that the protein folding is a process in two stages temporally separated: first, entropic forces are dominant and guide the chain into the native structure, and then, once in the native neighborhood, the HB (now protected from competition with the solvent), altogether with a more efficient structural specificity of the side chains (optimizing the van der Waals interactions), start the energetic stabilization of the protein.

Keywords: protein folding; thermal stability; Ts Kappa, molecular dynamics.

Lista de Figuras

Figura 1-1	Funil representativo de minimização de energia.	6
Figura 1-2	<i>Folds</i> conhecidos depositados no PDB	6
Figura 1-3	Degeneração de modelos em função da identidade	13
Figura 1-4	Estrutura e seqüência da Ts Kappa	16
Figura 1-5	Redução do espaço conformacional conferido pelas S-S	17
Figura 1-6	Esquema da formação de uma ligação S-S	18
Figura 1-7	Visualização 3D de uma ligação S-S	18
Figura 1-8	Experimento de Anfinsen, demonstrando a relação <i>folding</i> – S-S	19
Figura 3-1	Ts Kappa versão nativa e versão sem suas três S-S	35
Figura 3-2	RMSD e RG da Ts Kappa à T=293 durante 10 ns	40
Figura 3-3	RMSD e RG da Ts Kappa à T=348 durante 10 ns	41
Figura 3-4	RMSD e RG da Ts Kappa à T=398 durante 10 ns	42
Figura 3-5	RMSD e RG da Ts Kappa à T=448 durante 10 ns	44
Figura 3-6	RMSD e RG da Ts Kappa à T=498 durante 10 ns	45
Figura 3-7	Estruturas secundárias estabilizadas pelas LHs	46
Figura 3-8	Ângulos e distâncias atômicas restritivas às LHs	47
Figura 3-9	Exemplo de possíveis LHs em resíduos terminais da Ts Kappa	49
Figura 3-10	Exemplo de ocorrências competitivas	50
Figura 3-11	Exemplo de ocorrências não competitivas (envelope)	50
Figura 3-12	Ocorrência total de LHs no experimento de DM a T= 283K	52
Figura 3-13	Diferença de ocorrência de LHs entre as simulações com e sem S-S	53
Figura 3-14	Verificação da estabilidade da Ts Kappa nativa e variante (T=283K)	54
Figura 3-15	RMSD – Ts Kappa – 398K – Com S-S	58
Figura 3-16	RMSD e RG – Ts Kappa – 398K – sem S-S (desnaturação)	59
Figura 3-17	RMSD – Ts Kappa – 398K – Sem S-S (renaturação)	60
Figura 3-18	Raio de Giração – Ts Kappa – 398K – Sem S-S (renaturação)	60
Figura 3-19	Fotos de momentos-chave do vídeo da renaturação	61

Lista de Tabelas

Tabela 1-1	Análise dos principais modeladores por homologia <i>on-line</i>	11
Tabela 1-2	Valores energéticos de algumas ligações protéicas comuns	18
Tabela 3-1	Exemplo de separação da simulação por trechos de 2 ns	48
Tabela 3-2	Obtendo o envelope de um conjunto de ocorrências não competitivas	49

Lista de Abreviaturas

BLAST	<i>Basic Local Alignment Search Tool</i>	PDB	<i>Protein Data Bank</i>	PME	<i>Particle Mesh Ewald</i>
C	carbono	ps	pico segundos		
DM	Dinâmica Molecular	R	constant universal dos gases		
DNA	<i>deoxyribonucleic acid</i>	RG	Raio de Giração		
dt	<i>time step</i>	RMSD	<i>Root Mean Square Deviation</i>		
F força		RNA	<i>ribonucleic acid</i>		
FCFRP	Faculdade de Ciências Farmacêuticas de Ribeirão Preto	S-S	Ponte dissulfeto		
Fig.	figura	T	temperatura		
fs	femtosec	TSE	<i>Transition State</i>		
h	constante de Plank		<i>Ensemble</i>		
K	Kelvin	U	Estado desnaturado		
Kb	constante de Boltzmann	USP	Universidade de São Paulo		
Kcal	kilo calorias				
KJ	kilo Joules				
LH	Ligações de hidrogênio	Os 20 Aminoácidos:			
m	massa	Alanina	A	(ALA)	
MC	Monte Carlo	Arginina	R	(ARG)	
mol	moles	Asparagina	N	(ASN)	
N	Estado nativo	Aspartato	D	(ASP)	
N₀	número de Avogradro	Cisteína	C	(CYS)	
N	Nitrogenio	Fenilalanina	F	(PHE)	
nm	nanômetro	Glicina	G	(GLY)	
NMR	<i>Nuclear Magnetic Resonance</i>	Glutamato	E	(GLU)	
		Glutamina	Q	(GLN)	
NPT	<i>Ensemble NPT</i>	Histidina	H	(HIS)	
ns	nano segundos	Isoleucina	I	(ILE)	
O	oxigênio	Leucina	L	(LEU)	
PD	pontes dissulfeto	Lisina	K	(LYS)	
		Metionina	M	(MET)	
		Prolina	P	(PRO)	
		Serina	S	(SER)	
		Tirosina	Y	(TYR)	
		Treonina	T	(THR)	
		Triptofano	W	(TRP)	
		Valina	V	(VAL)	

Lista de Símbolos

α	letra grega alfa, encontrado em “ α -Hélice”
β	letra grega beta, encontrado em “Folha β ”
Å	Angstrom
Δ	letra grega delta, usualmente utilizada como símbolo de variação
θ	letra grega teta, representa ângulo protéico
ψ	letra grega psi, ângulo diedral
λ	letra grega lâmbda, comprimento de ondas

SUMÁRIO

Resumo	i
<i>Abstract</i>	ii
Lista de figuras	iii
Lista de tabelas	iv
Lista de Abreviaturas	v
Lista de Símbolos	vi

CAPÍTULO I

Modelagem molecular de Proteínas. Utilizando a restrição configuracional das Pontes Dissulfeto	01
1.0 Desvendando estruturas terciárias de proteínas: um desafio	01
1.1 Erro processo de Folding como causa de doenças	02
1.2 Biologia estrutural: estudo teórico-computacional.	03
1.2.1 PDB: o repositório de estruturas protéicas resolvidas	03
1.3 O problema do folding e da determinação da estrutura nativa de proteínas	04
1.4 Métodos automáticos de predição estrutural	07
1.4.1 Principais métodos computacionais por determinação da estrutura protéica	08
1.4.2 Modelagem Molecular por Homologia	09
1.4.3 Visão crítica dos principais modeladores protéicos por homologia on-line.	10
1.4.4 Modelagem por reconhecimento de <i>fold</i> (<i>Threading</i>)	13
1.4.5 Modelagem <i>ab initio</i> por dinâmica molecular	14
1.5 <i>Unfolding</i> (desnaturação) de proteínas	14
1.6 A toxina TS Kappa (pdb id: 1TSK) no presente trabalho.....	15
1.6.1 Pontes dissulfeto (S-S).....	16
1.6.2 A contribuição das S-S para o <i>folding</i>	16

CAPÍTULO II

2.0 Simulação computacional da conformação estrutural das proteínas.....	21
2.1 Dinâmica Molecular: uma visão geral	22
2.2 Fundamentos da Dinâmica Molecular.....	23
2.2 Campos de Força	25

2.0.2	Principais Campos de Força em Dinâmica Molecular de Proteínas: descrição e comparação de desempenho.	31
--------------	---	-----------

CAPÍTULO III

(Un)Folding em alta temperatura

3.0	Avaliando o papel das pontes dissulfeto	35
3.1	O processo de Unfolding.....	36
3.1.1	Cadeia original e cadeia sem pontes S-S.....	36
3.1.2	Parâmetros das simulações.....	37
3.1.3	Estabilidade conformacional	38
3.2	Dinâmica das Ligações de Hidrogênio na TS-Kappa	45
3.2.1	O papel das ligações de hidrogênio.....	45
3.2.2	A distribuição das LHs em função da presença de S-S.....	46
3.2.3	Variações das LHs que podem ocorrer em um par de resíduos	48
3.2.4	A distribuição das LHs em função da presença de S-S - Resultados	51
3.2.5	As LHs determinam a estabilidade da proteína - conclusões	48
3.3	Dinâmica Molecular em alta temperatura: concentrando em 398 K	56
3.3.1	Elongação da Simulação	57
3.3.2	<i>ReFolding</i> em alta temperatura: perspectivas	59
3.3	Conclusão e comentários finais	56

REFERÊNCIAS BIBLIOGRÁFICAS	63
----------------------------------	-----------

ANEXOS	71
--------------	-----------

Capítulo 1

Modelagem Molecular de Proteínas

1.0 - Desvendando estruturas terciárias de proteínas globulares: um desafio

Um biólogo vê as proteínas como biomoléculas, geralmente de grande peso atômico, traduzidas pela maquinaria ribossomal a partir do código genético, responsáveis pelas mais diversas funções metabólicas e estruturais de todos os seres vivos. Proteínas atuam na digestão celular e processos apoptóticos em forma de enzimas, comunicação celular (hormônios), respostas imunes (anticorpos), integração (citoesqueleto) e estruturação celular (proteínas fibrosas) etc. Essa variedade de funções é refletida pela variedade estrutural encontrada no universo das proteínas, já que a função de uma proteína é intrinsecamente ligada à sua conformação espacial.

No entanto, desde os experimentos termodinâmicos de Anfinsen (Anfinsen 1973) sabe-se que a conformação estrutural de uma proteína globular não é necessariamente determinada por processos biológicos, mas sim por processos físicos que dependem unicamente de sua sequência constituinte de aminoácidos e variáveis

ambientais tais como pH, temperatura, tipo e exposição de solvente etc. Assim, o estudo do processo que determina a estrutura terciária de uma proteína globular pode ser desenvolvido *in vitro*, e *in silico*, como um fenômeno físico-químico

Sob a perspectiva físico-química, proteínas são heteropolímeros que se diferenciam da maioria dos outros polímeros - sintéticos ou biológicos - por suas propriedades conformacionais peculiares: cada proteína dobra e enrola-se em uma única e precisa estrutura tridimensional, por meio de um processo denominado *folding*¹ de proteínas. E esta propriedade não é inconseqüente: existe uma fina sintonia entre a função e a estrutura das proteínas. Isto significa que, embora uma proteína seja identificada por sua seqüência de aminoácidos, ela cumprirá suas funções somente se estiver conformada espacialmente em sua estrutura nativa, a qual é alcançada em condições fisiológicas apropriadas.

1.1 – Erros no processo de *Folding* como causa de doenças

A importância do desenvolvimento técnico-científico no sentido de se acelerar a determinação rigorosa das estruturas protéicas e particularmente no entendimento das forças que governam o *folding*, é um passo imprescindível para o combate efetivo das chamadas doenças por mal enovelamento (*misfolding diseases*) (Ramos and Ferreira 2005; Chaudhuri and Paul 2006; Nakamura and Lipton 2009). O erro estrutural da proteína pode levar a graves problemas tanto pela perda de sua função quanto pela criação de um núcleo de agregação de outras proteínas (príons), causando acúmulos de proteínas antes funcionais em aglomerados altamente prejudiciais dependendo de sua localização no corpo humano. Doenças como Alzheimer, fibrose cística, encefalopatia espongiforme bovina (BSE - doença da “vaca-louca”) e muitas formas de câncer são exemplos de complicações originadas de proteínas com má formação terciária (*misfolded*); daí o interesse médico no problema: o conhecimento pleno dos mecanismos de folding levaria o desenvolvimento de fármacos a uma nova vanguarda.

¹ O termo “folding” tem sido traduzido em Português como “enovelamento” e também como “dobramento”. O primeiro é usado mais freqüente, mas também não traduz precisa e completamente a idéia contida no original, principalmente concernente às suas inflexões. Por isso em diversas instâncias neste texto, o termo em Inglês foi mantido.

1.2 - Biologia estrutural: estudo teórico-computacional.

Limitações intrínsecas da simulação do processo de *folding* decorrem basicamente do tamanho do espaço configuracional e da natureza da superfície (*landscape*) de energia livre. A composição destes fatores resulta em uma impossibilidade técnica computacional que não pode ser suplantada (pelo menos atualmente) sem o comprometimento do rigor nas representações das forças envolvidas e na constituição do sistema. Não se conhece ainda muito bem as implicações destas limitações, a não ser pelo estado ainda incipiente da área de predição teórica das estruturas de proteínas. Portanto, neste trabalho adotamos como estratégia uma metodologia que procura suplantá-la, de forma particular, a deficiência indicada. Assim escolhemos proteínas (i) de tamanho reduzido e (ii) com a presença de pontes dissulfeto (S-S), as quais podem ser representadas apropriadamente por certas neurotoxinas. Neste caso tais pontes podem ser usadas eficientemente como redutoras do espaço configuracional (fator que pode chegar à ordem de $N^{-3n/2}$, onde “n” é o número de pontes e “N” o número total de aminoácidos) (Gennes, 1985), a informação de quais cisteínas formam as pontes pode ser obtida como subproduto do seqüenciamento, ou por previsões baseadas em similaridades de seqüências.

Adotamos ainda condições iniciais otimizadas. Em geral, o estudo do processo de folding é conduzido tomando como base a cadeia distendida, (completamente desnaturada). Mas, é possível iniciar as simulações a partir de configurações muito mais próximas da estrutura nativa, tais quais as providas por serviços de modelagem molecular por homologia, descritos na seção 1.4.2.

Portanto, procurando satisfazer esses requisitos de factibilidade, elegemos uma proteína guia para nossa primeira parte do estudo, a qual é pequena e apresenta pontes dissulfeto (descrita na seção 1.6.1). O terceiro item (condições iniciais otimizadas) exigiu uma etapa de avaliação de um bom serviço de predição estrutural por homologia.

1.2.1 - PDB: o repositório de estruturas protéicas resolvidas

O número de seqüências protéicas já catalogadas é muito grande (3×10^6), mas com a introdução de novos genomas, como os de microorganismos marinhos, este número poderá facilmente triplicar (Yooseph, Sutton et al. 2007). De fato, graças às

técnicas eficientes de seqüenciamento, este número tende a crescer sistematicamente com a introdução dos genomas de espécimes de novos ambientes, como por exemplo, de amostras provenientes do solo ou de regiões mais profundas dos mares. Mas as técnicas experimentais de determinação estrutural (cristalografia por Raio X e ressonância magnética nuclear, essencialmente) são relativamente caras e lentas, e os métodos teóricos ainda estão em estágio de desenvolvimento: o sucesso da predição estrutural por comparação (*Swiss-Model*, por exemplo) é limitado, pois dependente da existência de um certo grau mínimo de identidade seqüencial de aminoácidos entre a proteína alvo e proteínas de estruturas conhecidas, e a predição por meio de primeiros princípios ainda carece não só de recursos tecnológicos mais avançados (apesar do tremendo desenvolvimento já alcançado nas últimas décadas), como também de progressos substanciais no entendimento dos fundamentos do processo de *folding*.

Por isso, atualmente somente cerca de uma entre 100 proteínas seqüenciadas tem sua estrutura depositada no *Protein Data Bank* (PDB- <http://www.pdb.org>), que possui atualmente aproximadamente 50.000 exemplares catalogados, muitas das quais apresentam variações sobre a mesma proteína. Contudo, apesar da taxa atual de crescimento de estruturas de proteínas depositadas anualmente no PDB ser grande, esta relação tende a piorar rapidamente com o contínuo seqüenciamento de novas proteínas. Portanto, este quadro faz com que a determinação/predição estrutural ganhe o *status* de problema emergencial.

1.3 - O problema do *folding* e da determinação da estrutura nativa de proteínas

A determinação experimental da estrutura terciária de uma proteína é, ainda hoje, um desafio se considerarmos processos em larga escala. Como já mencionado, o seqüenciamento de peptídeos caminha em passos mais largos que a resolução tridimensional de suas estruturas correspondentes, o que gera uma grande defasagem no que se diz respeito ao conhecimento seqüência-estrutura-função. O problema do *folding* de proteínas está se evidenciando cada vez mais como uma temática em que uma solução rápida não é uma realidade (Onuchic, LutheySchulten et al. 1997), e a maneira mais eficiente de tratá-lo é através da abordagem multidisciplinar, a qual inclui a biologia, química, física, matemática, computação etc. Gostaríamos de evidenciar que um dos destaques filosóficos do problema é o estímulo à multidisciplinaridade, a qual se

encontra gradualmente emergente na ciência moderna. Isso se deve ao nível molecular do problema, onde conceitos biológicos, físicos e químicos se encontram e se fundem.

A grande maioria das proteínas, não importando seu tamanho, é constituída somente por uma classe de moléculas orgânicas formando pequenas subunidades em sequência: os aminoácidos. Embora estes sejam encontrados em algumas centenas de tipos na natureza, apenas vinte foram essencialmente selecionadas evolutivamente para a constituição protéica.

Cyrus Levinthal (Levinthal, C. 1968) demonstrou que o elevado número de graus de liberdade de uma cadeia polipeptídica remete a um astronômico espaço conformacional de possíveis estruturas. Numa eventual busca aleatória pela única estrutura nativa, dentro o gigantesco espaço amostral, mesmo à altíssima taxa de uma tentativa por picossegundo, estima-se que o tempo característico consumido pelo processo seria maior que a idade atualmente creditada ao universo (aproximadamente 13,5 bilhões de anos!). No entanto, uma única geração da bactéria *Escherichia coli*, com todas as replicações protéicas necessárias, decorre no total de aproximadamente vinte minutos, o que prova que a experimentação aleatória e seleção por funcionalidade simplesmente não ocorre em seres vivos. Assim, o processo do *folding* seguiria certos caminhos privilegiados, porém tais rotas para o *folding* jamais foram mapeadas. Por isso, essa questão ficou conhecida por “paradoxo de Levinthal”.

Uma das metáforas comumente aceito para “vencer” o paradoxo de Levinthal é a da trajetória afunilada de energia livre (Leopold, Montal et al. 1992; Bryngelson, Onuchic et al. 1995; Onuchic, Socci et al. 1996), onde estados não nativos, que apresentam alta entropia, percorrem os vários estados intermediários em uma cadência para o estado nativo, até o ponto de convergência (a média de energia livre das estruturas freqüentemente obtidas em uma série de simulações tende a este ponto), quando o estado nativo é alcançado (Figura 1.1).

Figura 1-1. Representação do funil de energia: o caminho energético que as proteínas percorrem no processo de *folding* (Modificado de Fersht, 1999).

Porém ainda carecemos de uma descrição pormenorizada e geral dos mecanismos envolvidos no processo físico-químico de busca pela estrutura nativa, e avanços neste sentido são essenciais para o desenvolvimento de métodos para a determinação automática da estrutura terciária de proteínas.

O número de estruturas resolvidas experimentalmente e depositadas no PDB vem crescendo em um ritmo acelerado (embora insuficiente), enquanto o número de *folds* ou *motifs* (padrões estruturais) praticamente se estabilizou desde o ano de 2004 (Fig. 2). Isso é um indício positivo no campo da modelagem por homologia (próxima seção), já que a probabilidade de que a estrutura alvo pertença a um novo *fold* é mínima, o que garante uma maior confiança no sucesso do processo de alinhamento comparativo das sequências de duas proteínas.

Figura 1-2. O número de *folds* conhecidos depositados no *Protein Data Bank* se estabilizou praticamente em zero a partir de 2005. As barras mais escuras representam o incremento anual, enquanto as mais claras representam o número total. (Modificado de <http://www.pdb.org>)

1.4 - Métodos automáticos de predição estrutural.

O problema da determinação estrutural de proteínas tem sido abordado pelo uso de técnicas experimentais/computacionais como a cristalografia de raios-X, ressonância magnética nuclear (NMR), difrações de nêutrons ou simulações *in silico* (Fersht, 1999). Dentre todas as técnicas, as simulações computacionais são as mais criticadas pelas áreas mais tradicionais pelo fato de apoiar-se em bases ainda não totalmente sólidas, como a dos campos de força que ainda não reproduzem completa e adequadamente as interações inter-atômicas naturais. Esse trabalho é também estimulado por essa transitória carência existente na área computacional.

Os métodos automáticos de predição de estruturas protéicas têm despertado interesse devido ao baixo custo, rapidez na predição, e grande potencial de aplicação em diferentes áreas biológicas e da saúde, e em particular na indústria farmacêutica - além do atrativo acadêmico natural (desafio). Há basicamente dois tipos de abordagem:

(i) *métodos comparativos e métodos heurísticos (comparative protein modeling, protein threading, etc.)*: trata-se de algoritmos computacionais que se baseiam no "aprendizado" por meio de estruturas protéicas já conhecidas, e são considerados como restritos a "interpolações" (Karasawa, Tabuchi et al. 1993) e;

(ii) *métodos baseados em primeiros princípios*: técnicas de simulação que partem do conhecimento das interações energéticas efetivas em nível atômico e molecular, como Dinâmica Molecular (DM) e Monte Carlo (MC) (Karasawa, Tabuchi et al. 1993; Kolinski, Galazka et al. 1996; Ho and Dill 2006; Ho and Dill 2006; Irback and Mohanty 2006; Ulmschneider, Ulmschneider et al. 2006).

Embora em nenhuma das duas abordagens se pôde ainda lograr resultados gerais definitivos (Ghelis & Yon, 1982; Fasman, 1989; (Das, Matysiak et al. 2005; Rhee and Yang 2005), investimentos nos métodos de predição automática de estruturas tem se verificado de forma contínua, pois há grande confiança na possibilidade de uma conquista definitiva de métodos gerais e confiáveis para previsão automática de estruturas de macromoléculas *in silico*, embora, a rigor, evidências experimentais indicam que em muitos casos maquinarias biológicas são necessárias para conduzir a reação de *folding* adequadamente e na velocidade necessária (Hartl and Martin 1995; Hartl 1996; Welch and Brown 1996; Martin 1997; Sparrer, Rutkat et al. 1997; Hesterkamp and Bukau 1998; [Anon] 2007). Contudo, trata-se de um imperativo lógico que para parte das proteínas tais ajudas catalíticas não sejam compulsórias. Desse modo,

esta assertiva causal assegura a ocorrência espontânea do *folding* para algumas macromoléculas (proteínas pequenas são as candidatas mais prováveis). Outra evidência vem do fato que os *foldamers* (Hill, Mio et al. 2001) extrapolam as moléculas biológicas: certos polímeros não biológicos podem imitar proteínas no sentido de apresentarem a capacidade de se acomodar espontaneamente numa conformação exclusiva (única) e estável (Nelson, Saven et al. 1997; Pennisi 1997; Gellman 1998; Oh, Jeong et al. 2001).

Com os expressivos avanços teóricos e experimentais alcançados pela intensa atividade na área de predição estrutural, e o vigoroso ferramental introduzido por meio de computadores cada vez mais rápidos, é grande a expectativa de se conseguir progresso significativo - nas próximas décadas - na predição automática da estrutura de proteínas (a partir de sua seqüência de aminoácidos), e no entendimento pleno do processo do *folding* de macromoléculas. A suficiente generalidade do fenômeno (no sentido de seqüências nativas distintas possuírem a mesma habilidade para o *folding*) e conquistas recentes em diversas frentes de atuação (Kolinski, Skolnick et al. 1996; Dahiyat and Mayo 1997; Siew, Elofsson et al. 2000; Floudas, Fung et al. 2006) justificam o otimismo nessa direção.

1.4.1 – Principais métodos computacionais por determinação da estrutura protéica

Podemos dividir de maneira simples os métodos computacionais de determinação de estruturas protéicas em dois grandes grupos: a mais antiga e estudada, simulação tipo modelagem comparativa - homologia - e as relativamente mais novas técnicas, a saber, por primeiro princípios (ou *ab initio*; do latim “desde o começo”). Pela necessidade de modelos pré-trabalhados, dedicamos uma parte de nosso estudo à revisão teórica e prática do método por homologia, considerando alguns dos principais preditores de estrutura terciária em funcionamento on-line, com o objetivo de separar aquele que melhor se adequa às nossas necessidades. Além disso, procedemos com uma revisão bibliográfica sobre as origens, avanços e emprego da predição estrutural por comparação.

1.4.2 - Modelagem Molecular por Homologia

A origem da modelagem por homologia pode ser remetida a 1969, quando Browne *et al.* (1969) usando modelos em plástico utilizou a estrutura da lisozima e a comparou com a α -lactalbumina (as quais apresentam 39% de identidade sequencial), mantendo a estrutura comum e substituindo manualmente os aminoácidos diferentes na segunda (Petrey, Xiang et al. 2003; Wallner and Elofsson 2005). Essa experiência foi feita manualmente, e, como esperado, não apresenta boa precisão, mas tem valor pela introdução à idéia. Desde então, o método evoluiu muito e derivou-se principalmente em três grupos, os quais serão descritos a seguir.

O método de corpo rígido primeiramente promove o alinhamento da sequência de interesse com aquelas do banco de dados que apresentam maior identidade. Em seguida, divide-se estas sequências em pequenos blocos de estruturas secundárias (hélices, folhas β etc.), e de acordo com a maior identidade sequencial, associa-se o bloco mais apropriado na posição correspondente da sequência de interesse. Esse processo de substituição por blocos rígidos pré-determinados é feito para as sequências compartilhadas de aminoácidos, deixando as partes não alinhadas ao próximo passo do tratamento, que consiste na construção dos *loops*, orientação das cadeias laterais e das porções amino e carboxi-terminais. É neste passo que os vários programas se distinguem, já que cada um tem suas próprias regras para esse complemento por loops e outras partes não mudadas em relação ao banco de dados, as quais variam desde mudança por eventos evolucionários (Petrey, Xiang et al. 2003) até métodos de minimização por campo médio (Koehl and Delarue 1996). É interessante ressaltar que este mesmo método vem sendo empregado sistematicamente, desde os primeiros aplicativos de determinação estrutural protéica por homologia.

Existe também a abordagem do encaixe de segmentos, que consiste na utilização de um conjunto de posições atômicas derivadas à partir do alinhamento, o que orienta, por sua vez, a busca por seguimentos similares em um banco de dados de estruturas previamente conhecidas (Claessens, Vancutsem et al. 1989), os quais são selecionados de acordo com regras de energia e/ou geometria (Wallner and Elofsson 2005).

Por fim, a modelagem por homologia por restrições espaciais se apropria de uma série de restrições identificadas durante o alinhamento das sequências para, em seguida, dar início a um tratamento de minimização de violações estereoquímicas. Este tipo de modelagem é característica do programa Modeller (Sali and Blundell 1993).

1.4.3 - Visão crítica dos principais modeladores protéicos por homologia *on-line*.

Em primeira instância promovemos um breve estudo (o qual resultou em uma apresentação em forma de pôster no 6th *International Congress of Pharmaceutical Sciences* – CIFARP 2007) sobre algumas das principais ferramentas de predição de estrutura protéica *on-line* atuais. Isso foi necessário para o conhecimento desses mecanismos. Nesse estudo analisamos os servidores **SWISS MODEL** – *Biozentrum, University of Basel, Switzerland* (Schwede, Kopp et al. 2003); **FAMS** – *Kitasato University, Japan* (Ogata and Umeyama 2000); **3D-Jigsaw** - *Cancer Research, UK* (Bates, Kelley et al. 2001); e **CPHmodels** – *Technical University of Denmark* (Lund et al. 2002). O objetivo desse levantamento foi a investigação comparativa da performance e dos limites de predição dos programas, com isso tendo a base para eleger o melhor preditor por homologia em adequação à nossa proposta (ver seção 1.4.3).

A análise foi feita considerando-se o *Protein Data Bank* como o provedor de uma lista seleta de estruturas independentes: *i*) seqüências correspondentes foram enviadas à cada servidor e estruturas retornadas eram comparadas com àquelas conhecidas (controle) e, pelo cálculo do RMSD (*Root Mean Square Deviation*), o nível de consistência interna foi estimado; *ii*-) vários pontos distintos de mutação foram preparados e, novamente, submetidos aos modeladores, e as estruturas resultantes também foram comparadas pelo uso do RMSD.

É seguro destacar que nossos resultados tomam duas principais frentes. Primeiramente não existe perfeita consistência entre os diferentes serviços disponíveis, o que justifica a variedade deles. Segundo, alguns casos como simples mutações pontuais podem causar baixa consistência de resultado, até mesmo o fracasso em predizer a estrutura.

De acordo com a nossa experiência, compilamos uma tabela com os dados intuitivos e mais úteis a qualquer usuário que queira iniciar uma avaliação sobre os servidores (Tabela 1.1). A velocidade dos servidores foi em geral satisfatória com resposta, geralmente, em um ou poucos minutos, com exceção dos serviços **3D-Jigsaw** e **FAMS**, sendo este último o mais lento de todos, podendo chegar de algumas a muitas horas. O item “precisão“, como o próprio nome diz remete à qualidade da resposta obtida em relação à molécula nativa; esta medida foi quantificada pelo uso do cálculo do RMSD. A interface dos preditores é mais ou menos similarmente leve e de fácil

navegação entre eles todos, com exceção do **FAMS**, que recolhe as solicitações apenas via *e-mail*. A maioria apresenta campos de dados a serem preenchidos pelo usuário na própria *home page*, com opções de escolha para: **BLAST** *e-values* e possibilidades da escolha manual do molde (**SWISS-MODEL**); divisão interativa da sequência alvo em domínios, escolha de moldes e edição dos alinhamentos (**3D-Jigsaw**); *upload* da sequência **FASTA** (siglas de única letra de aminoácidos) à partir de um arquivo do disco rígido etc. Novamente, **FAMS** assume o último lugar e se mostra pobre na interatividade com o usuário.

O quanto as mutações interferem no alinhamento inicial constituiu-se num ponto importante a ser estudado, já que interfere diretamente na porcentagem de similaridade entre a sequência alvo e o modelo. Submetemos várias sequências mutadas aleatoriamente e algumas vezes seguindo padrões estereoquímicos de substituição (Henikoff and Henikoff 1992), obtendo resultados dos mais diversos. Como o estudo do **FAMS** se mostrou muito ineficiente, devido à lentidão do servidor, não computamos dados sobre este no que se diz respeito à tolerância a mutações. **3D-Jigsaw** e **CPH** mostraram-se bastante eficientes quanto ao alinhamento (não necessariamente corretos), enquanto **SWISS MODEL** se mostrou pouco mais exigente quanto ao nível de similaridade crítico à simulação.

Os três principais preditores se mostraram com velocidades diferentes de resposta inicial ao pedido do usuário: **SWISS MODEL** e **3D-Jigsaw** se mostraram mais dinâmicos e emitiam avisos rapidamente indicando que já estavam trabalhando no alinhamento. **CPH** se mostrou um pouco mais lento, mas nada muito relevante. Este último teste teve o propósito intrínseco de se investigar a velocidade da conexão do site, que não necessariamente é a mesma de sua capacidade de processamento.

	Swiss M.	3D-Jigsaw	CPH	FAMS
Velocidade	●●●●	●●	●●●●	●
Precisão	●●	●●	●	—
Interface	●●●●	●●●●	●●●●	●●
Mutações	●●	●●	●●	—

Tabela 1.1 Análise de alguns aspectos funcionais dos principais preditores de estrutura protéica estudados. A quantidade de pontos é crescente conforme o aumento de eficiência. Mais detalhes são descritos no texto.

Considerando-se que o número de estruturas resolvidas e depositadas no PDB tem aumentado exponencialmente, mesmo com o número de *folds* tendo permanecido estagnado desde 2005, nós apontamos o fato de que todos os preditores ainda falham em resolver consistentemente e com suficiente rigor muitos casos (obtivemos alguns RMSDs muito altos) de seqüências já conhecidas, o que deixa uma grande dúvida quanto à sua confiabilidade na determinação de estruturas de proteínas novas.

A literatura apresenta várias referências de comparação entre alguns programas de predição de estrutura protéica por métodos de homologia. Wallner e colaboradores (2005) demonstraram que alguns dos principais programas de predição por homologia apresentam atualmente distintas abordagens e, portanto, cada um possui uma especialidade intrínseca (modelagens de cadeias laterais, boa estéreo-química, confiabilidade, rapidez, diferentes taxas de deterioração do modelo *etc*). Isso demonstra que atualmente não há programas totalmente completos, o que deixa uma lacuna para melhoramentos futuros.

Foi demonstrado ainda, que, apesar de diferentes performances, os principais preditores atuais não apresentam, no geral, uma eficiência suficiente para suportar desenvolvimentos mais precisos como a modelagem de biomoléculas para uma síntese “*de novo*” em fármacos. Em vez disso, estas ferramentas devem ser vistas como um importante conjunto de métodos e técnicas em desenvolvimento que certamente muito contribuirão para a contribuição predição estrutural. Na figura 1.3 podemos observar que a maioria das vezes os programas baseiam-se em um modelo para a predição, o qual acaba sendo mais próximo da molécula nativa do que o produto final trabalhado. Isso significa que na maioria dos casos, os modeladores degeneram o modelo em moléculas de maior distância estrutural do alvo, fazendo justamente o caminho contrário esperado. Os dados abaixo indicam, porém que esse tipo de ferramenta deve ser considerada como uma boa possibilidade em casos de identidade seqüencial entre modelo e seqüência alvo abaixo de 55%, região na qual existe uma melhora nos resultados em geral.

Figura 1.3. O gráfico mostra a degeneração de modelos causada por alguns dos programas preditores por homologia em função da identidade da sequência. Observe em que todos os casos, a melhora é significativa apenas em identidades menores que 55%. (Modificado de [33]) Referências adicionais: Modeller – [37]; SegMod/ENCAD – [43]; nest – [34].

À luz destes dados, para a delimitação de nosso espaço estrutural consideramos como um bom procedimento a utilização conjunta das previsões estruturais dos diferentes programas de predição por homologia como informação (dados) inicial, visto que um dos nossos objetivos é a aplicação da DM como refinamento estrutural final.

1.4.4 - Modelagem por reconhecimento de *fold*s (*Threading*)²

O método de *threading* é muito similar à modelagem comparativa por homologia, já que também utiliza um banco de dados de estruturas tridimensionais previamente conhecidas, para executar uma comparação às ainda desconhecidas (Torda, Procter et al. 2004). Isso é feito através de uma função de marcação que avalia o ajuste de uma sequência à uma compilação de um certo *fold* e organiza os resultados em uma lista, a qual será analisada em busca do melhor molde a compor parte da estrutura alvo (Torda 2004). Neste método, Chang e colaboradores (Chang, Cieplak et al. 2001), emprestando técnicas da física estatística e redes neurais para o melhoramento do processo de *threading*, foi capaz de mostrar que certos parâmetros produzem uma medida quantitativa das tendências dos aminoácidos a serem internalizados ou expostos, ou mesmo darem preferência a um tipo determinado de estrutura secundária.

² Neste trabalho não utilizamos o método de *threading*; esta seção é objetiva, portanto, provemos apenas uma breve descrição da técnica.

1.4.5 - Modelagem *ab initio* por dinâmica molecular

Apesar de se apresentar em um estágio ainda prematuro, a modelagem *ab initio* se destaca pelo fato de independer de estruturas conhecidas para comparação configuracional. Este aspecto se emerge como ideal quando se reflete na quantidade de proteínas ainda por serem resolvidas, já que em muitos casos a máxima identidade de uma sequência com o melhor modelo se revela ainda muito baixa para uma comparação eficiente por homologia, ao mesmo tempo em que alinhamentos de altas identidades apresentam o problema da degeneração de modelos (Wallner and Elofsson 2005) – ver seção 1.4.3. Ademais a dinâmica molecular a DM é o único método computacional que provém uma análise dependente do tempo em um sistema em biologia molecular e, conseqüentemente pode ser implementado para se determinar o caminho energético específico da molécula, desde seu estado estirado ao nativo (Liwo, Khalili et al. 2005). De fato, mesmo a modelagem por homologia que geralmente apresenta certa eficiência quando se trata de predições de *motifs* de estruturas secundárias (folhas β e hélices), não existe procedimento computacional satisfatório para a previsão de *loops*, tendo que se recorrer à abordagem *ab initio* no momento de construção de alças e *loops*, que essencialmente consiste no cálculo das forças devidas às diversas interações atômicas no sistema (interações de *van der Waals*, potenciais hidrofóbicos e eletrostáticos etc.)

1.5 - *Unfolding* (desnaturação) de proteínas

O *unfolding* de proteínas, tratado com mais detalhes no Capítulo 2, é uma técnica utilizada como abordagem complementar no estudo do problema do *folding*. O processo consiste na desnaturação, geralmente por temperatura, de uma proteína específica, e subsequente observação do comportamento da estrutura geral da proteína, atentando-se às conformações de loops e extremidades, efeito sobre a capacidade enzimática, capacidade de *refolding* (reestruturação espontânea), etc.

Na parte inicial deste trabalho o processo de *unfolding in silico* foi massivamente realizada neste trabalho com o objetivo de ganharmos experiência e familiaridade com os métodos utilizados em DM; principalmente pela agilidade do processo e a conseqüente economia de tempo real e computacional. Assim, proteínas pequenas contendo pontes dissulfeto foram analisadas em diferentes temperaturas, mas por intervalos de tempo iguais para cada realização. Utilizamos referências da literatura

como guia de nossos passos e comparação de nossos resultados parciais. Em Seshasayee (Seshasayee, 2005), o *unfolding* por dinâmica molecular de uma proteína pequena nos forneceu uma boa base para o início de nossas atividades neste trabalho. O autor do trabalho referenciado utilizou a mini proteína *trp-Cage* (composta de apenas 20 resíduos) como objeto de estudo, enquanto que no presente trabalho utilizamos a proteína de 35 resíduos TS-Kappa (próxima seção). Seshasayee executou uma série de simulações com aumento de temperatura e estuda os efeitos sobre a estrutura da molécula, conseguindo delimitar três principais estágios do processo: inicialmente estado nativo (N), delimitação do estado de transição (TSE) e finalmente o constructo totalmente estendido e não-funcional, ou seja, *unfolded* (U).

1.6 A toxina TS Kappa (pdb id: 1TSK) no presente trabalho

A abordagem escolhida para lidar com o problema de predição estrutural é via *primeiros princípios*, não só por causa do interesse pelo processo de *folding* em si, mas também porque este é o único método teórico com potencial de fazer previsões estruturais de forma verdadeiramente independente: os demais métodos teóricos, como descritos acima, dependem essencialmente do conhecimento prévio de estruturas de proteínas com seqüências com alguma similaridade à da proteína alvo.

A escolha da proteína Ts Kappa se deve principalmente a duas razões: (i)- primeiramente por atender às exigências do projeto, quanto à redução do espaço das configurações: trata-se de uma proteína relativamente pequenas, e possui pontes S-S, condições que em conjunto constituem-se em eficientes fatores redutores do espaço das configurações espaciais da cadeia peptídica; e (ii)- também porque são proteínas de interesse farmacológico, tradicionalmente estudadas no Campus da USP de Ribeirão Preto, e em particular no Departamento de Física e Química da FCFRP-USP.

Existe cerca de 100 estruturas dessa classe depositadas no banco de dados de proteínas (PDB) com características similares à **Ts Kappa**, porém esta provém do veneno do escorpião amarelo brasileiro (*Tityus serrulatus*). Ela é suficientemente pequena, sendo constituída por 35 resíduos, sendo 20% de estruturas helicais (1 hélice de 7 resíduos) e 22% de folhas beta (3 folhas, com total 8 resíduos) e possui três pontes dissulfeto. É uma alfa toxina atuante nos canais iônicos de potássio (Blanc, Lecomte et al. 1997) (Figura 4-a).

Sua sequência é VVIGQRCYRSPDCYSACKKLVGKATGKCTNGRCDC, formando *cross-links* covalentes **S-S** entre C7-C28, C13-C33 e C17-C35 (Figura 4b). No *Protein Data Bank* a **Ts Kappa** é encontrada resolvida pela técnica de 2D-NMR, representada por um conjunto de 30 estruturas com variação de RMSD de até 4Å (Angstroms).

1.6.1 Pontes Dissulfeto (S-S)

Uma dos fatos mais interessantes da química de proteínas é que mesmo frente à sua enorme diversidade, são apenas 20 os aminoácidos para sua constituição (salvo o reino *Archea*). Cada qual com sua estrutura, estas biomoléculas apresentam características físico-químicas distintas entre si, tais como diferente eletronegatividade, polaridade, solubilidade, hidrofobicidade, capacidade de formação de ligações de hidrogênio, planaridade etc. Mas é na cisteína que iremos nos concentrar nesta seção,

pois este aminoácido é o responsável pela formação das únicas ligações covalentes conhecidas a conectarem *backbones* adjacentes: as pontes dissulfeto.

O atrativo deste *cross-link* para nosso trabalho é a minimização de espaço conformacional que este confere à proteína. Observe uma esquematização da delimitação estrutural que as S-S conferem a uma proteína com três ligações, como a Ts Kappa, na ilustração da Figura 1.5.

Figura 1.5. Demonstração esquemática da restrição do espaço conformacional que a introdução de pontes dissulfeto provê à estrutura. Observe que conforme o aumento de PD em uma proteína (da esquerda à direita) delimita cada vez mais sua variação em relação à estrutura nativa.

Dessa forma, a informação do posicionamento das PD por si só já confere uma vantagem considerável sobre uma estrutura puramente deformada. Isso é extremamente desejável, no presente contexto, pois nos coloca em um estágio de informação sobre a estrutura que talvez não conseguíssemos chegar por técnicas, hoje disponíveis, de primeiros princípios. No pior dos casos nos poupa tempo de simulação.

As pontes dissulfeto se formam espontaneamente em ambiente redutor, no entanto, biologicamente estas ligações são catalisadas pela classe enzimática isomerase. A reação se dá quando há proximidade suficiente entre dois grupos tiol (SH-HS) de cisteínas, os quais são respectivamente desprotonados e covalentemente conjugados; Figuras 1.6 e 1.7. A energia desta ligação é surpreendentemente grande (veja Tabela 1.2), aproximadamente dez vezes maior que a própria energia de ligação peptídica (Pauling, 1960).

Figura 1.6. Esquema da reação química para a formação de uma ponte dissulfeto.

Figura 1.7. Visualização de um modelo 3D de uma ponte dissulfeto. Imagem gerada através do programa Deep View- Swiss PDB Viewer (<http://spdbv.vital-it.ch/>)

Ligação	Energia (Kj/Mol)
S-S	213,1
C-C	83,1
C-O	84,0
C-N	69,7
Lig. Peptídica	21,0

Tabela 1.2. Valores das energias para as ligações mais comuns entre polipeptídeos (Pauling, 1960)

1.6.2 A contribuição das S-S para o *folding*

Estruturalmente, as ligações **S-S** desempenham papel fundamental na estabilização da proteína em casos de extrema temperatura (Capítulo 3) ou contra a degradação em ambientes oxidativos. Nas células eucarióticas, estas ligações são

formadas no lúmen do retículo endoplasmático rugoso, o qual tem seu redox oxidativo. Pelo fato do citoplasma se caracterizar como meio redutor, a maioria das proteínas dotadas de **S-S** se destinam ao meio extracelular. É por isso que encontramos essas ligações estabilizadoras mais comumente em substâncias de exportação, como toxinas e secreções.

Evidências experimentais e a existência de proteínas sem **S-S** determinam que estas não são imprescindíveis para o processo de *folding* de proteínas (Figura 1.8). Contudo, recentes estudos apontam que em alguns casos as pontes **S-S** podem ser formadas seletivamente durante o processo de *folding*, e que algumas regiões precisam ser estabilizadas por essas ligações, para garantir a estabilidade da região de estrutura secundária, para o início da próxima etapa (Zhou, Wang et al. 2005).

Figura 1.8. Experimentação de Anfinsen (Anfinsen 1973). Em cima: a proteína nativa é primeiramente reduzida, para a quebra das PD, depois desnaturada; depois é re-naturada e por último tem suas PD restauradas, retornando ao seu estado inicial em mais de 90% dos casos. Em baixo: a proteína nativa tem suas pontes dissulfeto desligadas e sua estrutura desnaturada; em seguida as pontes são novamente ligadas, mas dessa vez a re-naturação ocorre depois, o que determina proteínas aberrantes. Esse simples experimento sugere que as PD são formadas após o processo de folding.

Sabe-se que pontes dissulfeto podem tornar uma região protéica suficientemente rígida a ponto de esta perder a capacidade de acomodar-se formar ligações de

hidrogênio³ adicionais (Santiveri, Leon et al. 2008). Este é um dos assuntos que será abordado no Capítulo 3, quando a formação ligações de hidrogênio é analisada em função da presença e ausência das **S-S**, apresentando dados e resultados de em relação à estabilidade e consistência conformacional nas simulações em alta temperatura.

³ Ligações de Hidrogênio (antigamente conhecidas por pontes de hidrogênio) são uma das mais importantes dentre as chamadas ligações fracas (ao redor de 20 KJ/mol). Formam-se pela atração dipolo-dipolo entre grupos eletronegativos e um hidrogênio ligado à um nitrogênio, oxigênio ou fluoreto. Apesar de chamadas fracas, as LH em conjunto são de grande poder e fundamentais à estabilização principalmente de estruturas secundárias protéicas.

Capítulo 2

2.0 Simulação computacional da conformação estrutural das proteínas

A natureza não é fundamentada na simplicidade. Até mesmo as mais simples estruturas biológicas conhecidas pelo homem, os vírus, não podem ser considerados triviais, compreensíveis por meio de uma análise reducionista. Sabemos que os vírus mais simples são compostos apenas de uma carapaça protéica (capsídeo), a qual nada mais é do que a repetição de subunidades menores, interiorizando seu código genético (DNA ou RNA). Mas a aparente simplicidade acaba por aqui: ainda hoje não se sabe completamente os mecanismos de infecção de grande parte dessa categoria.

Atualmente é inviável sintetizar qualquer organismo conhecido pelo homem simplesmente pela manipulação direta de átomos, como um pequeno quebra cabeças, a partir de nenhum ponto de início pré-estabelecido, ou seja, do zero. Tomemos como exemplo a bactéria *Escherichia coli*. Conhecida do homem desde o fim do século XIX, e muito estudada até hoje como modelo padrão na maioria dos experimentos de biologia molecular envolvendo bactérias, a *E. coli* ainda é uma fonte de grande descobertas. A recente determinação de seu código genético (Blattner, Plunkett et al. 1997) trouxe mais questionamentos do que respostas.

Essa miríade de complexidade nos processos e organismos que nos cerca emerge com um problema em estudos com níveis de detalhamento atômico, como o tratado por esse trabalho. Por um lado é impraticável tratar de sistemas com parametrização total, é

necessário escolher as variáveis mais importantes a serem consideradas para cada caso. Por outro lado, essa abordagem por modelos minimalistas é muito comum e tem se mostrado eficiente ao longo dos anos. A seguir descrevemos alguns fundamentos de dinâmica molecular, técnica a qual faz uso constante de reducionismo para modelar átomos interagentes.

2.1 Dinâmica Molecular: uma visão geral

Nosso estudo teórico-computacional de estruturas protéicas é embasado na utilização da mecânica molecular (DM) para simular o sistema proteína-solvente em diferentes condições físico-químicas. Essa metodologia calcula basicamente a posição atômica de cada átomo constituinte do sistema em distintos e consecutivos instantes de tempo, separados por intervalos temporais da ordem de femto segundos ($1\text{fs} = 10^{-15}$ segundos). Este conjunto de coordenadas do núcleo de cada átomo do sistema, e sua evolução temporal, descreve a *trajetória* do sistema.

Geralmente, a DM se utiliza da descrição clássica newtoniana das forças envolvidas entre os constituintes do sistema. Por isso uma descrição pormenorizada das interações inter-atômicas, denominada *campo de forças*, está longe de ser suficiente para o tratamento rigoroso e geral dos sistemas moleculares. Assim, dada a complexidade e sutilezas interacionais envolvida nos problemas termodinâmicos e cinéticos das proteínas, surge naturalmente a questão sobre a necessidade de procedimentos mais rigorosas no trato das questões mais delicadas, como o do *folding* de proteínas. Um método amiúde mais preciso para a lida do problema é a mecânica quântica, a qual elevaria a mecânica molecular a um novo patamar, determinando as forças inter-atômicas diretamente por meio da consideração explícita dos elétrons dos orbitais relevantes (pelo menos), evitando (ou reduzindo) a utilização de potenciais energéticos efetivos. Esta abordagem mais rigorosa pode trazer dados valiosos ou desprezíveis, dependendo da situação, contudo invariavelmente a um enorme custo computacional. Por isso o emprego da mecânica quântica, no presente estágio do desenvolvimento tecnológico, fica restrito a sistemas suficientemente pequenos, e que exijam um maior refinamento nos tratamentos dinâmicos e termodinâmicos.

Por outro lado, a mecânica molecular clássica tem se desenvolvido a ponto de poder lidar satisfatoriamente em diversas situações com simulações de sistemas de tamanhos consideráveis, envolvendo até a ordem de milhões de átomos. Assim, muitas

vezes, não considerar o cálculo das interações eletrônicas é uma boa opção, senão a única. Ademais, para certos sistemas, efeitos quânticos podem ser desprezados sem grandes conseqüências, eliminando cálculos desnecessários e abatendo grande parte do custo computacional desnecessário. Essa simplificação permite simulações de maior duração, o que é fundamental ao estudo de certos processos, como é o caso do processo de *folding* de proteínas.

O fato de a mecânica molecular produzir resultados satisfatórios mesmo desprezando eventuais efeitos quânticos se deve a vários fatores, e a portabilidade dos campos de força é uma delas, isto é: parâmetros adequados para moléculas pequenas podem ser transpostos para grandes polímeros, da mesma forma que um pequeno número de casos é suficiente para descrever uma grande quantidade de sistemas diferentes. Outra importante propriedade da mecânica molecular clássica é que as energias potenciais são definidas em função das coordenadas moleculares, o que só é possível graças à aproximação de Born-Oppenheimer (Barnett and Landman 1993; Kubach and Rougeau 1998). Essa teoria se baseia na grande diferença entre massa m_e um elétron e a massa m_p de um próton ($m_p = 1836 m_e$), o que significa que os segundos podem ajustar-se quase que instantaneamente aos primeiros, sem influenciá-los.

Outro argumento usual a favor da descrição clássica das proteínas é que o comprimento de onda de deBrogliie térmico,

$$\Lambda = hN_0 / \sqrt{2\pi M_A RT}^4$$

para uma proteína de peso molecular M_A 30.000 Daltons, à temperatura ambiente, é da ordem de Λ 10^{-13} metros, o que é desprezível em relação aos comprimentos característicos da proteína (Ptitsyn, Finkelstein et al. 1999).

2.2 - Fundamentos da Dinâmica Molecular

A dinâmica molecular clássica consiste na aplicação das leis de Newton do movimento a um sistema molecular. Data as devidas condições iniciais, contendo as coordenadas espaciais e as velocidades de cada átomo, os parâmetros intensivos do

⁴ T é a temperatura; N_0 : numero de Avogadro; R: constante universal dos gases; e h é a constante de Plank.

sistema (temperatura, pressão) e o tipo de ensemble utilizado, na medida em que a simulação é executada, a evolução temporal (trajetória) de cada átomo, sob a influência dos demais átomos do sistema, é determinada por meio de algoritmos numéricos especiais. Os dados numéricos obtidos (coordenadas espaciais e velocidades basicamente) são decorrentes da solução numérica de um sistema de equações diferenciais acopladas, contendo em princípio três equações para cada átomo (uma equação para cada componente do espaço tridimensional) do sistema, que envolve. Operacionalmente, a cada instante se determina a força F (Anfinsen) sobre cada átomo, digamos átomo ‘ i ’, devido a todos os demais átomos (Anfinsen) do sistema, e por um intervalo de tempo δt muito pequeno o átomo ‘ i ’ se move então sobe esta força F mantida constante. A nova posição $\mathbf{r}_i$ do átomo ‘ i ’ assim como a e sua nova velocidade $\mathbf{v}_i$ são então *atualizadas* por meio da solução da equação (segunda lei de Newton):

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

Na prática, para cada partícula do sistema contendo um total de N átomos, é feito o cálculo das forças atuantes sobre a mesma, devido aos $N-1$ átomos restantes, a partir disso se obtém uma força vetorial resultante, e daí a aceleração instantânea que determinará a nova posição e velocidade do átomo posição no tempo imediatamente posterior $t+\delta t$. Esse processo é repetido durante toda a simulação, obtendo as novas posições e velocidades pra os instantes $t+2\delta t, t+3\delta t \dots t+n\delta t$.

Antes de se iniciar a integração das equações diferenciais, as velocidades iniciais dos átomos são geradas pela distribuição de Maxwell-Boltzmann, a uma temperatura pré-determinada, acordo com a equação abaixo:

$$P(v_{ix}) = \left(\frac{m_i}{2\pi K_B T} \right)^{1/2} \exp \left(-\frac{1}{2} \frac{m_i v_{ix}^2}{K_B T} \right)$$

onde o átomo 'i' de massa ' m_i ' tem a probabilidade $p(v_{ix})$ de ter a velocidade ' v_{ix} ' na direção 'x'. O parâmetro T é a temperatura absoluta e " k_B " é a constante de Boltzmann. As velocidades presentes no sistema são ajustadas de forma que o momento linear total do sistema seja igual a zero. Isso é feito pela soma das velocidades nas direções dos eixos x , y e z , cada qual é dividida pela massa total do sistema. Finalmente cada um desses três resultados são subtraídos das velocidades atômicas (Leach, 2001).

Ao fim de uma simulação de dinâmica molecular, a partir do conjunto relativamente simples de dados (sucessões de coordenadas tridimensionais) podemos proceder com várias análises. De uma simples informação de movimento entre átomos, é possível a obtenção de grandezas como temperatura e pressão do sistema; momento, aceleração e velocidade das partículas; forças de torção planar em moléculas; rotações, variações e estiramento de ligações atômicas e seus ângulos; formação e destruição de ligações (como geração e quebra de ligações de hidrogênio); informações sobre a termodinâmica do sistema (como os diversos tipos de energia: eletrostática (*coulombiana*), Lennard-Jones, cinética *etc*); dados sobre flexibilidade e deformação da estrutura protéica, bem como sua compactação; enfim uma infinidade de parâmetros.

Todavia, a trajetória só é obtida de forma razoavelmente realista por utilização de potenciais contínuos (cada partícula terá sua força alterada a cada mudança de posição própria ou de seus vizinhos). Contudo, a adição desse potencial contínuo acarreta um "pareamento" de dependência entre moléculas, originando um problema relacionado com o efeito de muitos corpos. É então que a impossibilidade de resolução analítica nessa situação demanda que a integração pelas equações *newtonianas* sejam através de um método diferencial com elementos finitos, realizado por algoritmos específicos de integração. Existem atualmente vários tipos de integradores, todos podem ser aproximados a uma série de Taylor (para as variáveis de posição e cinética), são os principais Verlet (Verlet 1967) , Leap-Frog (Hockney, 1970) , Velocity-Verlet (Swope, Andersen et al. 1982) e Beeman (Beeman 1976).

2.3 Campos de Força

As interações que ocorrem entre os átomos constituintes da cadeia polipeptídica se dão em níveis energéticos que cobrem três ordens de grandeza: de frações a centenas de Kcal/mol. Assim, os vários tipos de interações energéticas relevantes têm sido

classificados, e o conjunto sistematizado delas se denomina por Campo de Força, e é ele quem governa toda a dinâmica do sistema, produzindo a trajetória de cada átomo do sistema. Uma comparação entre os principais campos de força disponíveis atualmente é feita mais adiante, com o objetivo de ilustrar o quanto a participação deste mecanismo é importante para o resultado final da simulação.

Em geral um campo de força é formado por dois componentes distintos. Primeiro, um conjunto de funções potencial, dependente somente das posições relativas de cada átomo (energia potencial), é utilizado para se gerar as forças sobre cada átomo, através do gradiente do potencial, isto é,

$$F(\mathbf{r}_i) = -\nabla_i U(\{\mathbf{r}_j\}) = -\left\{ \frac{\partial U(\{\mathbf{r}_j\})}{\partial x_i} \hat{\mathbf{x}} + \frac{\partial U(\{\mathbf{r}_j\})}{\partial y_i} \hat{\mathbf{y}} + \frac{\partial U(\{\mathbf{r}_j\})}{\partial z_i} \hat{\mathbf{z}} \right\},$$

e, segundo, um conjunto de parâmetros numéricas é determinado por meio de ajustes em sistemas moleculares mais simples, para então ser *portado* para sistemas mais completas. A esse processo se denomina *parametrização* do sistema. Neste trabalho vamos apresentar alguns detalhes somente do primeiro componente dos campos de força

Os potenciais são basicamente divididas em (1) não ligadas, (2) ligadas e (3) especiais. Interações não ligadas são computadas baseando-se em uma lista de átomos vizinhos não ligados, dentro de um certo raio de corte. As ligadas são representadas por ligações covalentes (potencial de estiramento), dobramento angular, diedros próprios e impróprios. Ligações especiais remetem a potenciais especiais, os quais são usados para impor restrições no movimento do sistema, tanto para evitar desvios desastrosos ou para incluir conhecimento de dados experimentais. Incluem restrições na posição, nos ângulos, nas distâncias e nas orientações. No entanto estes potenciais restritivos não são implementados nos softwares que iremos utilizar neste trabalho (GROMACS – veja anexo I), portanto não serão aprofundados neste trabalho.

São funções não-ligadas:

Lennard Jones (LJ): essa interação se dá pela equação 2.1 abaixo, onde C_s dependem dos tipos de átomos, retirados de uma matriz de parâmetros LJ.

$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6}$$

(2.1)

Modificado de Spoel et. al, 2006.

Potencial de Buckingham: é uma função mais sofisticada que a anterior, pois seu termo de repulsão é mais flexível e realista (eq. 2.2). No entanto essa característica reflete em maior custo computacional.

$$V_{bh}(r_{ij}) = A_{ij} \exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^6}$$

(2.2)

Modificado de Spoel et. al, 2006.

Interação de Coulomb: gera uma curva um tanto quanto diferenciada em relação às outras duas anteriores (figura abaixo).

$$V_c(r_{ij}) = f \frac{q_i q_j}{\epsilon_r r_{ij}}$$

(2.3)

onde $f = \frac{1}{4\pi\epsilon_0} = 138.935\,485 \text{ (m/F)}$

Modificado de Spoel et. al, 2006.

Por sua vez, são funções ligadas:

(I) Potencial de Estiramento: a flexibilidade da ligação entre átomos i e j é representada por um potencial harmônico, como ilustrado na figura, representado no gráfico e descrito na equação abaixo. Onde k_{ij} e b_{ij} dependem do par de átomos (i,j).

$$V_b(r_{ij}) = \frac{1}{2} k_{ij}^b (r_{ij} - b_{ij})^2$$

(2.4)

Modificado de Spoel et. al, 2006.

(II) Potencial Angular Harmônico: a vibração do ângulo formado pelos átomos i, j e k é representada por um potencial harmônico no ângulo θ_{ijk} :

$$V_a(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta(\theta_{ijk} - \theta_{ijk}^0)^2 \quad (2.5)$$

(III) Diedros Próprios: neste caso existe a opção de escolha de interação entre a função normal e a baseada na expansão em potência de $\cos\psi$ (chamada de Potencial de Ryckaert-Bellemans).

(a) para a função normal, temos abaixo:

$$V_d(\phi_{ijkl}) = k_\phi(1 + \cos(n\phi - \phi_s)) \quad (2.6)$$

Modificado de Spoel et. al, 2006.

(b) analogamente, para a função Potencial de Ryckaert-Bellemans, temos abaixo:

$$V_{rb}(\phi_{ijkl}) = \sum_{n=0}^5 C_n (\cos(\psi))^n,$$

(2.7)

C_0	9.28	C_2	-13.12	C_4	26.24
C_1	12.16	C_3	-3.06	C_5	-31.5

Constantes para o Potencial de Ryckaert-Bellemans em (kJ mol⁻¹)

Modificado de Spoel et. al, 2006.

(IV) Diedros impróprios: Diedros impróprios agem na manutenção da planaridade de grupos já planos (como anéis aromáticos) e prevenir o espelhamento de uma molécula. Abaixo a equação e seu gráfico representativo:

$$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_{\xi} (\xi_{ijkl} - \xi_0)^2$$

(2.8)

Modificado de Spoel et. al, 2006.

A cadeia polipeptídica é descrita com resolução atômica, com cada átomo “*i*” sendo representado por uma massa pontual *m_i*. O “campo de forças” finalmente, é obtido relacionando-se as funções apresentadas até este ponto, da seguinte maneira:

$$V(\{r_{ij}\}_N) = [V_{\text{comprimento}} + V_{\text{angulos}} + V_{\text{torções}}]_{L_Covalen} + V_{L_hidrog} + V_{nao_Ligadas}$$

A força f_i sobre cada átomo “ i ” dentre os N átomos da macromolécula, por sua vez é dada por

$$f_i = -\sum_j \nabla_i V(\{r_{ij}\}_N).$$

Cada termo da Equação 2.9 é função das coordenadas $\{rij\}$ dos átomos da macromolécula. Os potenciais são separados em três grupos principais: ligações covalentes (variações nos comprimentos das ligações, nos ângulos entre cada três átomos consecutivos, e torções entre uma ligação e outras ligações vizinhas); ligações de hidrogênio; e interações que não constituem “ligações”, como as do tipo Coloumbiana e van der Waals. Detalhes sobre as funções potencias e dos parâmetros envolvidos na simulação de proteínas podem ser obtidos, por exemplo, no site: <http://www.gromacs.org>.

Com isso fechamos a parte descritiva das funções de potencial e partimos (próxima seção) para a comparação de alguns dos campos de força mais populares atualmente.

2.0.2 Principais Campos de Força em Dinâmica Molecular de Proteínas: descrição e comparação de desempenho.

No o final da década de 70, as simulações moleculares surgiram (Scott, Hunenberger et al. 1999) e vinham engatinhando como uma técnica utópica que jamais conseguiria simular grandes moléculas, muito menos em períodos de tempo longos. Foi então que o *boom* da informática (início dos anos 90) permitiu o avanço de técnicas computacionais cada vez mais complexas, as quais vêm sendo estudadas e desenvolvidas no campo da ciência. É neste contexto que tanto a dinâmica molecular quanto Monte Carlo vêm recebendo novas tecnologias para serem utilizadas em vários campos acadêmicos. O campo de força computacional chegou com a necessidade (e permissibilidade) da simulação de moléculas *in silico*, e desde então é objeto sistemático de estudo.

Campos de força vêm sendo modificados para estudo de carboidratos (Kony, Damm et al. 2002), proteínas (Xu, Luo et al. 2007), produtos naturais (Kahn and Bruice 2002) etc., comparados com dados experimentais e validados (Wensink,

Hoffmann et al. 2003; Suardiaz, Perez et al. 2006), o que demonstra sua versatilidade e importância.

Contudo, um problema surge da seguinte máxima: se a natureza apresenta apenas um campo de força, (pelo menos na Terra; excluem-se singularidades, buracos negros etc) e a informática apresenta vários, ou (1) apenas um está certo ou (2) nenhum está certo. De qualquer forma, embora indique prospecto pouco agradável, o estudo comparativo de diferentes campos de força se mostra indispensável ao avanço da área.

Aqui vamos fazer uma pequena revisão de algumas comparações encontradas na literatura (Kaminski and Jorgensen 1996; Chen, Martin et al. 1998; Imberty, Gruza et al. 1998; Villamanan, Chamorro et al. 2006), entre *AMBER96*, *CHARMM22*, *GROMOS 43A1* e *OPLS-AA*. Todos, com exceção de *TraPPE*, foram parametrizados para estudos em proteínas, com grupos funcionais dos 20 aminoácidos. A falta de estudos específicos de comparação de campos de força para dinâmicas moleculares de proteínas dificulta muito uma comparação neste nível, no entanto, algumas conclusões puderam ser tiradas.

(I) AMBER96

AMBER96 utiliza o potencial de (12-6) *Lennard Jones* mais cargas pontuais para átomos soltos ou distantes; para átomos diferentes é utilizado ainda as regras de mistura de *Lorentz–Berthelot* (Hess, Kutzner et al. 2008). Interações intra-moleculares apresentam dinâmicas do potencial tipo harmônico tanto para estiramento de ligações quanto para torções angulares e ângulos diedrais do tipo próprio. Átomos conectados por três ligações apresentam tratamento misto: são regidos por *Lennard Jones* e potencial de Coulomb.

(II) CHARMM22

Este campo de força utiliza as mesmas regras de potenciais não ligados que AMBER96: potencial de (12-6) *Lennard Jones* mais cargas pontuais para átomos soltos ou distantes e as regras de mistura de *Lorentz–Berthelot*. As diferenças são que as interações intra-moleculares apresentam dinâmicas do potencial tipo harmônico tanto para estiramento de ligações quanto para torções angulares e ângulos diedrais do tipo impróprio, além de termos não escalados do tipo Coulomb. Torções angulares do tipo

impróprias são utilizadas em casos especiais com fins de restrição de planaridade de resíduos aromáticos.

CHARMM22 é considerado um dos melhores campos de forças segundo estudo de estados vapor-líquido [12], bem como bastante popular para essa finalidade (Imberty, Gruza et al. 1998), só ficando atrás do específico TraPPE (Martin 2006).

(III) GROMOS 43A1

GROMOS 43A1 é um campo de força baseado na unificação dos átomos, juntando a maioria dos hidrogênios aos carbonos correspondentes. É parametrizado para usar as mesmas formas funcionais que AMBER96, sem a regra geral de mistura de *Lennard Jones*. As relações intra-moleculares são diferenciadas. Flexibilidade das ligações é produzida pelo potencial de estiramento de quarta potência, enquanto ângulos se retorcem com torções do tipo próprias e os diedrais são tratados pela série de cossenos, *Lennard Jones* e termos de Coulomb não escalados. Podemos citar como diferencial do GROMOS 43A1 o fato de que é o único dentre os citados que aplica uma única série de cossenos diedral ao longo de qualquer par central de átomos, efeito um tanto quanto desprezível para moléculas grandes como proteínas. Para simulações de líquido, GROMOS 43A1 apresenta resultados muito emparelhados com AMBER96 e OPLS-AA.

(IV) OPLS-AA (*Optimized Potentials for Liquid Simulations – All Atoms*)

Este popular campo de força apresenta muito em comum com AMBER96 e CHARMM22: potencial de (12-6) *Lennard Jones* mais cargas pontuais para átomos soltos ou distantes; e para átomos diferentes é utilizado ainda mistura de regras geométricas. Interações intra-moleculares apresentam dinâmicas do potencial tipo harmônico tanto para estiramento de ligações quanto para torções angulares e ângulos diedrais do tipo próprio. Átomos conectados por três ligações também apresentam tratamento misto, porém em escala distinta: são regidos por *Lennard Jones* e potencial de Coulomb, ambos em escala de 0,5. Torções angulares do tipo impróprias são utilizadas em certos átomos do tipo sp² para restringir a planaridade de resíduos aromáticos.

Como popular campo de força em estudos protéicos, o OPLS-AA se sai bem na maioria dos parâmetros, no entanto apenas dados sobre simulações com alcanos foram encontradas nessa resenha (Martin and Siepmann 1998).

Capítulo 3

(Un)Folding em alta temperatura

3.0 Avaliando o papel das pontes dissulfeto

O problema do *folding* de pequenas proteínas globulares é visto neste trabalho como um processo que possui necessariamente dois estágios, separados temporalmente. No estado *unfolded*, o sistema proteína-solvente é instável, e é nessa condição que se dá a primeira fase do processo, isto é, a *busca* pela estrutura nativa. Neste estágio inicial de busca, forças entrópicas são dominantes, e durante a evolução temporal conformações estruturalmente mais prováveis (aquelas com muitos contatos entre resíduos próximos ao longo da cadeia) se sucedem. Eventualmente, a conformação da cadeia se aproxima suficientemente da estrutura nativa, quando o segundo estágio do processo se inicie: com um número suficientemente grande de LH inter-cadeia protegias do solvente e de resíduos otimamente compactados (fortalecendo as interações de van der Waals) dá-se início a fase da estabilização energética do sistema, com a cadeia se compactando sistematicamente na conformação globular.

Neste capítulo serão apresentados resultados de simulações de DM referente às duas etapas do *folding* acima citadas. Assim, por um lado o papel dos *cross-links* **S-S** e das LH serão analisadas comparativamente, como ingredientes da estabilidade protéica. E por outro lado, utilizando-se do artifício de elevar a temperatura, as forças hidrofóbicas serão enfatizadas (pois estas são as únicas que não participam da competição cadeia-solvente) para analisarmos o seu papel no processo de busca pela estrutura nativa.

Assim, neste capítulo nossos resultados foram organizados em três tópicos principais. No primeiro, relata-se os resultados das simulações exploratórias iniciais, cujo principal objetivo foi o entendimento do comportamento conformacional da proteína em questão durante o processo denominado *unfolding*, isto é, o quanto e como a proteína se afasta de sua estrutura nativa à medida em que a temperatura é incrementada. No segundo tópico analisamos o papel dos *cross-links S-S* e das LH com respeito à estabilidade conformacional. E, finalmente, no terceiro tópico, analisamos a possibilidade e o significado do *folding* em altas temperaturas.

3.1 - O processo de *unfolding* -

O senso comum leva a crer que quanto mais conexões energéticas, como as ligações **S-S** estiverem presentes nas macromoléculas, maior resistência a intempéries físico-químicas estas serão capazes de suportar sem perda da integridade estrutural e, portanto, sua função. E isso pode ser considerado válido, porém somente em casos específicos. Nossas análises demonstram que tão importante quanto a quantidade de **S-S** é seu posicionamento, o qual pode até vir alterar a flexibilidade da molécula (seção 3.3.2).

Todos os experimentos computacionais aqui apresentados, se valeram do pacote de simulação e análise de dinâmica molecular GROMACS v3.3.1, que foi introduzido no Capítulo 2.

3.1.1 - Cadeia original e cadeia sem pontes S-S

Duas preparações da Ts Kappa (**1TSK**), uma contendo as três **S-S** originais, e outra sem ligações **S-S**, Figura 3.1, foram submetidas a uma bateria de simulações independentes. Os experimentos são todos idênticos, excluindo-se particularidades explícitas, como a diferenciação por temperatura e presença de **S-S**.

Primeiramente escolhemos uma janela de tempo plausível (o que sempre é condicionada pelas capacidades computacionais disponíveis), o qual nos permitisse obter um mínimo de conhecimento da evolução da molécula em diferentes temperaturas sem saturar nosso equipamento. Dessa forma, as primeiras trajetórias foram geradas com duração de 10 ns (nano segundos) cada.

A TS Kappa foi originalmente determinada por NMR, à temperatura de 283K, que por essa razão tornou-se uma temperatura de referência no nosso trabalho. Em seguida, aumentamos sistematicamente os valores da temperatura, em simulações independentes, de modo a varrer da melhor maneira o espaço de busca para esse problema. Nessa seção nos concentramos a descrever essa série de experimentos e apresentar seus resultados.

Figura 3.1. Duas preparações da Ts Kappa. A estrutura da esquerda refere-se à nativa, determinada por NMR. A outra refere-se ao resultado da simulação durante 10 ns do sistema cadeia-solvente no qual as três pontes **S-S** da cadeia foram removidas. Observe a diferença entre as hélices; isso é explicado pela reordenação das ligações de hidrogênio.

submetidas a um *ensemble* NPT (Número de partículas, Pressão e Temperatura constantes) e realizadas e analisadas através do pacote GROMACS versão 3.3.1 (para maiores informações e tutorial de instalação e simulação, veja o **anexo I**). Prezando pela uniformidade experimental, todos os experimentos neste trabalho compartilham de um mesmo conjunto básico de parâmetros. No eventual caso de parametrização em temperaturas diferentes, estas mudanças serão devidamente indicadas no texto. Segue abaixo alguns dos os principais parâmetros de base:

- **Campo de força OPLS/AA-L** (Jorgensen and Tirado-Rives 1998): em poucas palavras, é um dos mais utilizados para a desnaturação térmica (*unfolding*) de proteínas (observe a seção 2.0.2);

- **Água explícita SPC3:** todos os eventos se deram dentro de uma caixa virtual cúbica de aresta 5.39524 nm, a qual foi preenchida com 4974 moléculas de água Simple Point Charge (Alper and Levy 1989), cuja integridade é confirmada por sua vasta referência na literatura (Baez and Clancy 1995; Bagchi, Balasubramanian et al. 1997; Crozier, Henclerson et al. 2001; Eriksson, Jacobi et al. 2008);
- **Termostatos de temperatura e pressão de Berendsen:** conjunto de algoritmos desenvolvidos por Berendsen (Berendsen, Postma et al. 1984), os quais modelam a temperatura e pressão no sistema;
- **Equilíbrio iônico:** a fim de neutralizar as cargas implícitas geradas em pH 7,0 pelas porções N(+) e C(-) terminais, quatro lisinas (+), três argininas (+) e dois aspartatos (-), foram inseridos **oito** íons CL^- para contrabalancear as cargas positivas e três íons Na^+ para contrabalancear as negativas ;
- **PME (*Particle Mesh Ewald*):** é o método de soma de Ewald para eletrostática de longa distância, mais eficiente (ganho em desempenho) que a soma de Ewald clássica;
- **dt (time steps):** intervalo de tempo para a integração numérica do sistema de 0,002 ns;
- **Raios de corte:** para as interações eletrostáticas e de van der Waals, utilizamos o raio de corte com valor de 1,0 nm.

3.1.3 - Estabilidade conformacional

Duas grandezas são utilizadas na análise da estabilidade conformacional em função da temperatura: a compactação da cadeia polipeptídica, dada por seu Raio de Giração (RG), e seu distanciamento conformacional da estrutura nativa ao longo do

tempo, medido pela desvio quadrático médio RMSD (a seção 2.1 do **anexo I** explica como se obtém essa medida).

É comum uma estrutura determinada experimentalmente apresentar certo relaxamento estrutural quando submetida à simulação por DM. Por esse motivo, procedemos com uma simulação à sua temperatura de determinação experimental da estrutura, a saber, $T = 283\text{ K}$. De fato observamos uma pequena variação em relação à sua conformação espacial inicial, bem como em seu nível de compactação. Ambos as grandezas flutuam sistemática e levemente ao redor 0.2 nm para RMSD e ao redor 0.8 nm para RG, mas não apresentam valores tais que indiquem qualquer alteração detectável na estrutura nativa.

Ao aumentarmos a temperatura em 10 K, praticamente nada se altera. As variações continuam dentro daquilo observado na Fig. 3.2, a não ser por uma pequena perturbação que ocorre por volta do instante 4000 ps, no caso com **S-S**, para o RG. Isso pode ser explicado como uma acomodação/contração espontânea, a qual acaba se dissipando a medida que o tempo evolui.

Figura 3.2 Variação conformacional da TS Kappa durante o experimento de dinâmica molecular em água explícita na temperatura de **293K**. O gráfico superior representa a simulação da molécula com as **S-S**, enquanto o inferior representa a variação sem **S-S**. O RMSD (linha inferior, mais espessa) indica o quanto a estrutura se distanciou de seu estado inicial (em nanômetros), nesse caso o nativo. O raio de giração (RG), o qual é representado pela linha superior (mais clara e fina), mede o quão grande foi a compactação (valores baixos) ou expansão (valores altos) da cadeia protéica (também em nanômetros).

Pela pequena variação apresentado no RG e RMSD, nas duas temperaturas, decidimos aumentar a temperatura por um salto de $\Delta T = 50K$. É que neste ponto pensamos em lançar mão de uma lógica que lembra vagamente a de Newton-Raphson para resolução de raízes de funções: no caso de uma variação muito grande nos resultados para o novo valor da temperatura (no caso 348 K), escolheríamos agora um valor intermediário entre as duas temperaturas, e no caso oposto, dobraríamos o incremento do caso anterior.

Todavia, os valores ainda continuaram muito próximos dos anteriores, Fig. 3.3, apesar de se mostrarem um pouco mais elevados e com flutuações maiores no modelo sem **S-S**. Este fato é interessante porque mesmo com essa temperatura bastante elevada, ambos os casos (com e se **S-S**) se saem praticamente igualmente estáveis. Isso *poderia* significar que as pontes dissulfeto não teriam papel fundamental na estabilização térmica da estrutura como um todo, no entanto uma discussão mais detalhada acerca do tema será feita logo à frente (seção 3.2). Embora as simulações mostradas tenham sido realizadas para um tempo relativamente curto de 10 ns, para verificar se a estabilidade térmica foi alcançada, simulações mais longas (20, 40 e 100 ns) de alguns casos foram executadas, confirmando os resultados.

Figura 3.3. Variação conformacional da TS Kappa durante o experimento de dinâmica molecular em água explícita na temperatura de **348K**. O gráfico superior representa a simulação da molécula com as **S-S**, enquanto o inferior representa a variação sem **S-S**. O RMSD (linha inferior, mais espessa) indica o quanto a estrutura se distanciou de seu estado inicial (em nanômetros), nesse caso o nativo. O raio de giração (RG), o qual é representado pela linha superior (mais clara e fina), mede o quão grande foi a compactação (valores baixos) ou expansão (valores altos) da cadeia protéica (também em nanômetros).

Assim, mais um incremento de 50 K foi feito à temperatura do sistema, o que revelou fatos interessantes: os no início da simulação os valores do RMSD e RG são praticamente os mesmos (RMSD $\sim 0,2$ nm e RG $\sim 0,8$ nm), porém com flutuações em ambos os casos (com e sem S-S) mais acentuadas (Fig. 3.4).

Figura 3.4. Variação conformacional da TS Kappa durante o experimento de dinâmica molecular em água explícita na temperatura de **398K**. O gráfico superior representa a simulação da molécula com as S-S, enquanto o inferior representa a variação sem S-S. O RMSD (linha inferior, mais espessa) indica o quanto a estrutura se distanciou de seu estado inicial (em nanômetros), nesse caso o nativo. O raio de giração (RG), o qual é representado pela linha superior (mais clara e fina), mede o quão grande foi a compactação (valores baixos) ou expansão (valores altos) da cadeia protéica (também em nanômetros).

Continuando, aumentamos a temperatura uma vez mais. Agora com o sistema a 448K, observamos que em ambos os casos, houve sinais claros dos efeitos da alta temperatura agindo sobre sua integridade estrutural. Primeiro notamos que o RMSD aumenta rapidamente embora com compactação ainda mais ou menos constante. Os dados da Figura 3.5 que com certeza contrariam o que se poderia esperar, é o fato de que no caso da simulação da cadeia com as pontes **S-S**, ocorre uma mudança conformacional forte e precoce (≈ 300 ps), em relação ao caso sem **S-S**. Note também que por volta de 7000 ps, o RMSD do modelo com **S-S** começa a aumentar, indicando desnaturação protéica, o que seria esperado. Esse comportamento pode ser explicado observando que quando vínculos estruturais são removidos, quando as ligações **S-S** são extintas, novas LH podem se estabelecer com um novo re-arranjo estrutural fino. Nesse momento, a alta temperatura causa flexionamento de uma parte da estrutura, a qual afasta-se crescentemente de sua forma globular. À medida que isso ocorre, aumentam também as forças geradas pelo efeito hidrofóbico, as quais podem atrair essa porção desgarrada crescentemente e, nesse caso em particular, acabar sobrepujando as colisões do solvente, levando essa região de volta à posição inicial. Tudo indica, porém, que mais um estágio de aumento na temperatura será suficiente para que a energia cinética do solvente seja mais que suficiente para sobrepor qualquer restituição conformacional que possa vir a ocorrer.

Figura 3.5. Variação conformacional da TS Kappa durante o experimento de dinâmica molecular em água explícita na temperatura de **448K**. O gráfico superior representa a simulação da molécula com as **S-S**, enquanto o inferior representa a variação sem **S-S**. O RMSD (linha inferior, mais espessa) indica o quanto a estrutura se distanciou de seu estado inicial (em nanômetros), nesse caso o nativo. O raio de giração (RG), o qual é representado pela linha superior (mais clara e fina), mede o quão grande foi a compactação (valores baixos) ou expansão (valores altos) da cadeia protéica (também em nanômetros).

Com a instância da verificação da validade de nossa explicação sobre a reafirmação da estrutura ocorrido no estágio anterior, um último experimento, com aumento de temperatura, foi realizado. Agora sofrendo efeito das colisões das moléculas de água a 498 K, como esperado, as proteínas vão se deformando cada vez mais e começam, portanto, a se desnaturar (Fig. 3.6). Observe que nessa temperatura extrema, agora sim, a **S-S** (por se tratar de uma ligação covalente) se destaca como força estabilizadora de grande eficiência.

Figura 3.6. Variação conformacional da TS Kappa durante o experimento de dinâmica molecular em água explícita na temperatura de **498K**. O gráfico superior representa a simulação da molécula com as **S-S**, enquanto o inferior representa a variação sem **S-S**. O RMSD (linha inferior, mais espessa) indica o quanto a estrutura se distanciou de seu estado inicial (em nanômetros), nesse caso o nativo. O raio de giração (RG), o qual é representado pela linha superior (mais clara e fina), mede o quão grande foi a compactação (valores baixos) ou expansão (valores altos) da cadeia protéica (também em nanômetros).

Essa bateria de experimentos iniciais, com e sem as pontes **S-S**, nos foi de grande valor para a verificação de certas hipóteses sobre o processo de *folding*, como ficará mais evidentes no decorrer dos tópicos seguintes.

3.2 - Dinâmica das Ligações de Hidrogênio na TS-Kappa

Qual o efeito da remoção das pontes dissulfeto para a estabilidade protéica? Esta é a pergunta cuja resposta tínhamos como alvo quando planejamos as simulações de dinâmica molecular da proteína Ts-Kappa.

Nesta etapa, realizamos duas simulações, ambas com duração total de 20 ns. A primeira simulação tratou da cadeia íntegra, e a segunda considerou a mesma cadeia polipeptídica, porém agora com a remoção das três pontes **S-S** (a saber Cis7-Cis28, Cis13-Cis33 e Cis17-Cis35) que interconectam diferentes regiões da proteína original.

3.2.1 - O papel das Ligações de Hidrogênio

Como visto anteriormente, as pontes **S-S** são ligações covalentes fruto da oxidação de duas cisteínas adjacentes na estrutura terciária da proteína. Pela natureza altamente energética dessas ligações, é natural inferir que quanto maior sua quantidade,

maior será a estabilidade estrutural da proteína. Por outro lado, encontramos LHs com muita abundância como força de estabilização exclusivas em vários motifs secundários, como em alfa-hélices e folhas beta (Fig. 3.7). Estas duas ligações químicas possuem características diferentes e propósitos diferentes, no entanto é possível encontrar as LHs em posições que contribuem para a estabilidade geral da proteína, e não apenas na estrutura secundária.

Devido à competição com solvente pelas LH o que reduz na variação marginal da energia livre entre o estado nativo e o unfolded (equivalente à energia de algumas poucas LH) o papel das LH na estabilidade é um pouco controverso (ref-1), mas nessa seção mostramos alguns resultados de como se relacionam as LHs em função das S-Ss, e como a estabilidade da proteína pode ser afetada pelo deslocamento dessas ligações fracas entre os resíduos.

Figura 3.7. (a) As alfa hélices são estabilizadas em forma helicoidal pelas LHs (mostradas em listras vermelhas) entre os C=O com os grupos N-H; (b) as folhas betas são consolidadas também por LHs, porém em disposição que gera uma folha pregueada em planos angulados sucessivos (c).

3.2.2 - A distribuição das LHs em função da presença das S-S

Assim, as trajetórias obtidas, isto é, o registro da evolução espaço-temporal de todos os átomos da proteína e do solvente, foram analisadas com foco nas LHs. A ferramenta básica utilizada é o programa *g_hbond*, o qual é parte integrante do pacote

computacional previamente citado. O *g_hbond* analisa distâncias e ângulos relativos em que se dispõem os átomos propensos à formação de LHs (H, O e N) (Fig. 3.8), interpretando essas informações e determinando onde e quando ocorreram as ligações de hidrogênio durante a simulação. No entanto essas informações são apresentadas de uma forma pouco prática para nosso fim, ou seja, é apresentada a contagem de todos os momentos em que possa ter ocorrido uma ligação de hidrogênio em algum lugar da proteína. Com o objetivo de tornar relativos (em percentuais) estes valores absolutos, utilizamos um script (**anexo II**) gentilmente fornecido a nós pelo Dr. Carlos Alessandro Fuzo, que na época finalizava seu doutoramento pelo departamento de Química da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (Universidade de São Paulo). Dessa forma obtivemos na forma percentual a ocorrência de todas as LHs durante a simulação, onde valores próximos de 0% indicam a inexistência de ocorrência de LHs durante a simulação, enquanto os próximos de 100% significam que os resíduos estiveram ligados pelas LHs o tempo todo.

Figura 3.8. Condições a serem satisfeitas para a formação de ligações de hidrogênio. A distância entre o oxigênio e o hidrogênio é dada para esse caso específico, no entanto esse valor varia de acordo com os átomos envolvidos, bem como temperatura e pressão. Apesar da relativamente grande permissibilidade de abertura angular representada, as maiores energias de ligação se dão quando COH e OHN se aproximam de 180° (valores angulares extraídos de Baker & Hubbard, 1984).

Contudo, para simulações suficientemente grandes como este caso, o valor de ocorrência total não é suficiente para descrever satisfatoriamente de que forma se deu a ligação entre os aminoácidos, já que estatisticamente dois conjuntos totalmente diferentes de dados podem possuir exatamente a mesma média. Isso quer dizer que em uma simulação de 20ns, um par qualquer que tem a ocorrência total igual a 50%, pode

ter permanecido ligado 100% por metade da simulação, 50% durante toda a simulação, 25% no início e 75% a partir da metade etc. Por isso, utilizamos uma implementação muito útil do script do Dr. Fuzo, a qual além de calcular a ocorrência total das LHs, separa a trajetória de cada par em trechos, retornando a porcentagem de ligação parcial para cada um deles. Observe na tabela abaixo (Tab. 3.1) um exemplo de separação por trechos⁵.

intervalo (ps)	2000	4000	6000	8000	10000	12000	14000	16000	18000	20000	Média
13CYS - 25THR	18,4	40,4	59	64,3	66	58,2	22,1	6,9	0,9	2	33,82

Tabela 3.1. Exemplo da separação dos dados em médias por trechos. Observe que a média (33,82) não representa bem esse caso, já que os valores flutuam muito, com um pico de ocorrência de 66% no intervalo 8001-10000 e um vale de 0,9 em 16001-18000. A posse dessas informações é essencial para, entre outros fatores, a verificação da estabilidade da ligação durante todo o período. O par aqui exemplificado pode ser encontrado também nas figuras 3 e 4, sob o índice 17.

3.2.3 Variações das LHs que podem ocorrem em um par de resíduos

É importante ressaltar que ao longo das simulações é natural que ocorram posicionamentos eventuais e transitórios entre distintos grupamentos atômicos dos resíduos envolvidos ou do *backbone*, os quais embora possam satisfazer as condições de distância e angulação para qualificação de uma LH, durante uma pequena fração de tempo, não devem ser considerados como ligações estáveis e, portanto, válidas. No entanto, estes eventos são de também computados, o que pode acabar gerando uma porcentagem inválida de ocorrência de LH ao fim da simulação. Para evitar isso, utilizamos um filtro de dados, trabalhando apenas com os pares que apresentaram em pelo menos um trecho da trajetória uma porcentagem de ocorrência suficientemente consistente, a qual assegure que se trate de uma LH efetiva. Após a separação apenas de valores maiores ou iguais a 20%, os dados obtidos foram reduzidos de um total de 162 ocorrências para 45 (com **S-S**) e de 186 para 46 (sem **S-S**).

O motivo para o filtro de 20% se aplicar aos trechos da trajetória e não à média de ocorrência de cada par de resíduos é utilizado para resolvermos o problema da emergência de múltiplas LHs. Os pares de aminoácidos, principalmente os de cadeia lateral longa como a lisina e a arginina, podem estabelecer mais que uma LH,

⁵ Lista de abreviações no início do trabalho contém as referências dos códigos de três letras dos vinte aminoácidos essenciais.

simultânea ou não (Fig. 3.9). Ligações não simultâneas (Fig. 3.10) se dão quando uma LH impede a formação de outra, por excesso de doadores competindo por apenas um aceptor, ou vice versa. Ligações simultâneas (Fig. 3.11) podem ser independentes ou não, e se dão quando ambos doadores e aceptores estão presentes em mais que uma unidade. Logo, em um dado momento na simulação, na importa quantas LHs existam, mas sim qual é a mais freqüente, porque é esta que entra como representante no envelope das curvas, o qual seleciona os valores necessários para o novo e correto valor de ocorrência total de LH no par como um todo. A tabela abaixo (Tab. 3.2) sumariza todo esse processo, evidenciando o caso da Valina 1 – Cisteína 35.

Tempo (ps)	2000	4000	6000	8000	10000	12000	14000	16000	18000	20000	Média
(I) 01VAL - 35CYS	0,4	15,8	99,7	92,2	99,6	99,5	100	99,2	99,5	99,2	80,51
(II) 01VAL - 35CYS	99,6	83,4	0	7,8	0	0	0	0	0	0,6	19,14
(III) 01VAL - 35CYS	92,8	96	94	96,8	93,6	92,1	98,2	96,1	95,8	95,9	95,13
Envelope	99,6	96	99,7	96,8	99,6	99,5	100	99,2	99,5	99,2	98,91

Tabela 3.2. Obtendo o conjunto de ocorrências resultante a partir dos pontos máximos (em negrito) de cada trecho. A média dos pontos máximos, a qual representa a verdadeira média de ligação dos resíduos, é maior do que aquela que seria escolhida, caso optássemos por selecionar apenas o valor contido na última coluna a direita (par **III**). Aqui fica claro como o filtro de 20% por intervalos se aplica de modo favorável. O conjunto **(II)** possui uma média muito menor que as demais, abaixo inclusive do valor crítico de 20%. No entanto é exatamente este conjunto o responsável pelo maior valor no tempo 0-2000 ps, o qual figura na linha dos pontos máximos.

Figura 3.9. a) visão geral da face da TS-Kappa, a qual é detalhada (área demarcada) em (b), mostrando apenas algumas das várias possibilidades para formação de LHs entre o oxigênio da cisteína 35 com os grupos NH da valina 1 e cisteína 23, bem como o grupo NH da cisteína 35 com o oxigênio da valina 1.

Figura 3.10. Distribuição da ocorrência de LH, em intervalos de 2ns, da simulação da TS Kappa com as três pontes dissulfeto originais. Aqui ilustramos como a LH pode variar entre átomos diferentes em um mesmo par de resíduos, exemplificado aqui pela Valina 01 ligada à Cisteína 35: a curva **I** corresponde a LH entre o grupo NH3 da valina com o grupo O1 da cisteína. O índice 3 desse grupo, assim como os índices dos demais grupos, indica apenas o número de ordem do hidrogênio no resíduo, conforme convenção. As demais curvas, **II**, **III**, e **IV**, representam os outros casos observados. Note que a ligação **I** se mantém estável até o instante correspondente a 100 ns, quando o grupo NH3 é inclinado a ligar-se ao O2 da cisteína, aumentando a representação da curva **II** na porcentagem total de ligação deste par. No entanto, instantes mais tarde o O2 da cisteína troca a LH corrente para a nova ligação com o NH2 da valina (curva **III**), a qual, por sua vez, é rapidamente substituída pela última, e provavelmente também transitória, ligação observada, onde o NH2 da valina retorna ao grupo de oxigênio original do começo da simulação, o O1 da cisteína (curva **IV**).

Figura 3.11. Envelope das ocorrências: curva resultante da conexão dos pontos instantâneos máximos ao final de cada intervalo de 2 ns, para o par Lisina 23 – Cisteína 35. Cada curva representa uma distinta ligação que pode ocorrer competitivamente (em diferentes instantes de tempo) ou cooperativamente (ao mesmo tempo). Tal atividade na troca das LH entre grupos atômicos distintos do mesmo par de resíduo ocorre tanto na TS-Kappa com a preservação das S-S ou sem as mesmas.

Como nosso objetivo aqui é estudar a estabilidade da molécula em relação a ocorrência de LHs, necessitamos encontrar apenas o quanto cada par de resíduos permaneceu ligado durante a simulação e, por isso, tratamos indiscriminadamente os grupos aceptores e doadores, assim ambos são contabilizados da mesma forma para a criação do já citado envelope das curvas.

Após a condensação das ocorrências por envelopamento das curvas das LHs múltiplas nos pares, temos dados ainda mais refinados, de forma que os pares significantes agora caem de 45 para 21 (com **S-S**) e de 46 para 23 (sem **S-S**). Nessa etapa, os dados precisam ser juntados de maneira a permitirem finalmente uma comparação.

Apesar de tratarmos em ambas as simulações de uma mesma proteína, o fato de uma delas possuir as pontes dissulfeto desligadas provoca a formação de alguns pares de LHs únicos. Então, como exemplo, o par treonina 25 – glicina 26 ocorre em média 32% na simulação da TS Kappa com **S-S**, e inexistente na outra, tendo sido, portanto, adicionado como um par válido, mas com uma ocorrência de 0%. Após relacionarmos as simulações dessa forma, os pares de LHs em sua forma final somam 30 para ambos os casos.

3.2.4 – A distribuição das LHs determinadas pelas pontes S-S - Resultados

Nesse momento, com todos os pares entre as duas simulações tendo correspondência, foi possível criar o gráfico mostrado na figura 3.12. Esse gráfico apresenta as porcentagens de ocorrência de todos os pares que passaram por todo o processo de filtragem (30 pares). As barras escuras correspondem aos valores da simulação com **S-S**, e, por sua vez, as barras claras correspondem aos valores da simulação sem **S-S**.

O primeiro resultado importante que encontramos é a confirmação de que as simulações realmente diferenciam-se tanto em relação ao número de pares que fazem LHs, quanto aos valores de ocorrência destes. No entanto, o que estamos realmente interessados é em focar exatamente nas diferenças que surgiram, e por isso rearranjamos os pares de forma a evidenciar aqueles que mais variam de um experimento para outro. Dessa forma, criamos um terceiro conjunto de dados, os quais são resultados da diferença entre a ocorrência dos pares na simulação da TS Kappa com **S-S** e a correspondente sem **S-S** (Fig. 3.13).

Observando essas diferenças entre as simulações, detectamos uma alta simetria entre os dados, o que indica que o número geral de LHs não varia entre as simulações, ou seja, a remoção das pontes dissulfeto na TS Kappa não tem um efeito direto sobre o número de LHs. Entretanto, como vemos na Fig. 3.12, existem vários pares que se encontram muito mais intensos em um experimento do que em outro.

Figura 3.12. Porcentagem de ocorrência de ligações de hidrogênio pelos pares correspondentes durante o tempo total da simulação (20 ns, T= 283K). As barras mais escuras representam os valores referentes à simulação da proteína íntegra, ou seja, com as três pontes dissulfeto nativas intactas. Por sua vez, as barras cinza-claro representam o caso obtido na simulação com a remoção das pontes dissulfeto. As referências dos pares de aminoácidos apresentados no gráfico (números de 1 a 30 na abscissa) se encontram ao lado direito da figura, sendo que os números que imediatamente precedem a abreviação dos resíduos correspondem ao seu posicionamento na proteína.

Figura 3.13. Diferença entre as ocorrências de ligações de hidrogênio entre a simulação da proteína nativa e a proteína sem S-S. Valores positivos relacionam-se à maior frequência de formação de LH ao par correspondente quando a proteína está intacta do que em relação ao segundo caso. Valores negativos representam os pares que mais aparecem quando a TS-Kappa não apresenta as ligações dissulfeto, enquanto valores próximos de zero indicam que não existiu mudança significativa na frequência de LHs entre as duas simulações.

Entre os dez pares mais representativos, no caso sem pontes **S-S**, sete envolvem pelo menos uma cisteína (a qual é responsável pela formação das **S-S**), diferentemente do caso com a presença das **S-S**, que apresenta apenas quatro pares com pelo menos uma cisteína entre os dez mais intensos. Isso se deve à restrição espacial imposta pelas pontes **S-S** aos átomos dos resíduos adjacentes, os quais ficam impedidos de ajustarem-se nos ângulos e distâncias apropriadas para a formação de LHs. Isso fica evidente no trabalho de Zhou et al. (2005), quando foi conseguida uma elongação espontânea de alfas hélices (estabilizada exclusivamente por LHs) em proteínas que originalmente continham pontes **S-S** e tiveram sua estrutura mutada para evitar tais ligações. A remoção das **S-Ss** permite então o livre movimento dos grupos aceptores e doadores das cisteínas e de resíduos próximos. Isso é significativo do ponto de vista da estabilidade da proteína, pois a remoção de três ligações covalentes (portanto, fortes) pode levar a crer em uma desestabilização do empacotamento da estrutura, e não é isso que observamos.

De acordo com medições de deformação estrutural da proteína ao longo da simulação, ambas as estruturas permanecem igualmente estáveis, como ilustra a Figura 3.14. Então a manutenção de estabilidade, somada ao aumento de LHs entre as cisteínas no caso sem **S-S**, aparece como forte evidência da compensação da remoção das **S-Ss** pelo rearranjo das LHs.

Figura 3.14 Apesar de um pequeno aumento na deformação da proteína sem **S-S** no início da simulação, verificamos que não houve diferenças estruturais marcantes entre as versões da TS Kappa. Essa equivalência mostra que na ausência das PD, o rearranjo de LHs nas diferentes regiões da proteína pode compensar, nas condições das simulações, é suficiente para manter a proteína em sua forma nativa.

O par 10 (cisteína 07 – glicina 31), conecta a base da terceira fita beta ao primeiro *coil* da TS Kappa, foi aquele que mais aumentou na simulação sem **S-S**, chegando a uma participação de 92%. O interessante é que na presença das pontes dissulfeto, a cisteína 07 já não tem capacidade de formar LHs, sendo por isso que nessas condições, a glicina 31 aparece alternativamente ligada à glutamina 05 praticamente na mesma alta ocorrência.

Existe também a relevância dos pares 01 e 02, que envolvem a valina 01. Esse resíduo se mantém conectado à cisteína 35 em uma alta taxa (por volta de 97%) tanto na

presença quanto na ausência das **S-S**, porém nesse último caso, a valina 01 aparece com uma ligação adicional à asparagina 34, o que leva a um aumento da força de ligação entre os extremos da proteína em 77%. O terceiro lugar de nossa lista, o de número 8 (Glutamina 05 – Cisteína 33), também contribui bastante para a estabilização dessa região quando não existem as **S-S**.

Quanto aos pares que pouco variam, eles podem ser encontrados ao redor da posição mediana do eixo x da figura 7. É curioso chamar a atenção para os pares que apresentaram menor variação entre todos os outros (2, 4, 29 e 30), pois conferindo seus valores absolutos através da figura 6, podemos constatar que estes são pares conservados de ocorrência alta e provavelmente de grande importância à estrutura nativa da TS Kappa. Por sua vez, alguns pares que também não variaram muito (como o 26 e 22), apresentaram valores absolutos baixos em ambas as simulações, o que mostra a importância de utilizar ambas as figuras em questão para tirarmos conclusões mais refinadas com relação à ocorrência das LHs.

3.2.5 – As LHs determinam a estabilidade da proteína - conclusões

Estes resultados nos permitem fazer alguns comentários conclusivos, mas antes devemos diferenciar o conceito de estabilidade, como a variação da energia livre entre o estado nativo e o estado *unfolded*, do conceito mais geral de estabilidade conformacional, no sentido de preservação da forma e tamanho do glóbulo protéico. Assim, uma proteína poderia perder sua estrutura nativa, sem necessariamente mudar sua conformação, como ilustrado na Figura 3.5.

As simulações com e sem pontes **S-S** da TS-kappa sugerem que as LHs são as responsáveis pela estabilidade da proteína, sendo que o papel das pontes **S-S** seria a de proteção da proteína contra um afastamento demasiado da estrutura nativa, no caso de grandes flutuações locais de grandezas intensivas (temperatura, pH), o que lhe conferiria uma chance maior de *refolding*. De fato, a remoção das pontes **S-S** apenas determinaram um re-arranjo das LHs, mantendo a essência de sua constituição estrutural nativa, mesmo com variações importantes da temperatura. Evidentemente, por mais pequena que seja, uma mudança configuracional pode alterar ou mesmo anular a função da proteína. Assim, embora o conjunto das pontes **S-S** e das LH possam ser determinantes da função da proteína, é de muita significância o fato revelado pelas simulações aqui desenvolvidas, que a estrutura nativa pode não sofrer alterações

significativas com a remoção das pontes **S-S**. Coisa que certamente não ocorre com o contrario, isto é, a estrutura nativa é totalmente perdida sem a LH, mesmo que as pontes **S-S** sejam mantidas.

3.3 Dinâmica Molecular em alta temperatura: concentrando em 398 K

Para a elongação da simulação e concentração de nossas análises, escolhemos então trabalhar com a temperatura de 398 K. Como descrito na seção anterior, pudemos observar que, nessa temperatura, a proteína é capaz de se deformar de forma suave, ao mesmo tempo em que ainda é levemente capaz de devolver regiões de sua molécula em estado próximo ao nativo quando de uma desestabilização.

Com devidas ressalvas, esse valor de temperatura na solvatação da Ts Kappa pode ser considerado como um fator catalisador à movimentação molecular.

Consideremos mecanismos de otimização por elevação da temperatura utilizados pelo corpo humano como dispositivo de defesa. A febre é um estado em que o corpo força o aumento sua temperatura, acelerando os movimentos moleculares e, dessa forma, a velocidade de reação das enzimas e anticorpos. No entanto, não é razoável associar a temperatura com a qual estamos trabalhando aqui (398K, i.e. 125°C) com um estado febril humano. Não se tem notícia de nenhum metazoário que seja capaz de sobreviver a essa temperatura. Porém, de fato existem organismos, obviamente compostos de maquinaria protéica, integrantes do reino Archea, como os do gênero *Sulfolobus*, os quais habitam desde águas acidificadas em fontes termais a 87 °C até chaminés vulcânicas no leito oceânico sob pressão enorme e temperaturas por volta de 105 °C (Lesk 2003). Isso significa que existem proteínas que se mantêm estáveis e funcionais em temperaturas muito maiores que o senso comum pode levar as pessoas a crer. De forma surpreendente, essas proteínas *termoestáveis* não apresentam muitas diferenças quanto às suas parentes mais abundantes, são formadas pelos mesmos aminoácidos, sendo as únicas diferenças o maior empacotamento e o número aumentado de pontes salinas (Lesk 2003).

Como sabemos, a função protéica está intrinsecamente relacionada à estrutura da mesma, além de que a atividade enzimática é baseada na flexibilidade molecular. No entanto, um período febril alongado é fatal ao indivíduo, pois apesar da temperatura ter uma influência positiva ao metabolismo protéico, esta também gera ocasionalmente pontos de desnaturação estrutural, o que leva à destruição da integridade da molécula. É

por essa coerência que pensamos em nossas simulações: a alta temperatura nos permite obter resultados em menor tempo computacional, porém irá fatalmente danificar a molécula a longo prazo.

A parte mais técnica da simulação também deve ser investigada cuidadosamente, afinal como pode uma proteína solvatada em água líquida a 398K? De fato, as simulações por desnaturação térmica geralmente são feitas a temperaturas superiores a 498K (Scheraga, Khalili et al. 2007), com o intuito de instituir uma rápida desnaturação, gerando uma grande quantidade de dados em tempos curtos. Nessas condições é imperativa a parametrização sob o *ensemble* NPT, onde a pressão constante impede a água de vaporizar-se e a solvatação não ocorreria da maneira correta.

3.3.1 Elongação da Simulação

Um fator relevante, do ponto de vista computacional, determinado pela alta temperatura, é que nestas condições o sistema pode “escapar” muito mais facilmente dos mínimos energéticos locais (Fig. 1.1), e assim o espaço configuracional pode ser mais rapidamente varrido, mesmo que por períodos transitórios. Essencialmente, o que ocorre, é que a estabilidade da proteína, medida pela diferença da energia livre ($F = E - TS$) entre o estado nativo e o *unfolded*, $\Delta F = F(N) - F(U)$, é reduzida e se aproxima sistematicamente de kT à medida em que T aumenta.

Este fato tem efeitos interessantes em relação às forças que induzem a cadeia para a estrutura nativa e às forças que estabilizam a proteína. Com suficiente aumento da agitação térmica, as LH se formam e se desfazem com maior frequência, o que põe em evidência a competição entre a formação de PH intra-cadeia e cadeia-solvente, destruindo a estabilidade do glóbulo. Por outro lado, como é conhecido pelo da termodinâmica da energia livre de transferência de materiais hidrofóbicos da água para meios apolares, embora a variação da entalpia e da entropia sejam significativas com a temperatura, a sua composição na variação da energia livre, $\Delta F = \Delta H - \Delta TS$, quase não sofre mudanças no intervalo entre 273 e 350K (Lee 1991). Assim, o efeito do aumento da temperatura sobre o processo de busca pela nativa, induzido pelas forças entrópicas (hidrofóbicas), é praticamente muito baixo.

Munidos deste cenário, realizamos simulações da TS-Kappa sem as pontes **S-S** na temperatura de 398K, com tempos de duração crescentes (Figs. 3.2 a 3.6). Iniciamos

com uma janela de tempo modesta de 20ns, mas já nesta primeira etapa o comportamento do RMSD mostrou uma tendência de crescimento sistemático, o que nos instigou à alongação da simulação até os 50 ns.

Concluimos que a flexibilidade que a falta de **S-S** confere, é apropriada para acessar vários níveis energéticos da proteína, o que não ocorre no modelo com **S-S** intactas, devido as restrições causadas pelas mesmas (observe a baixa variação em uma simulação ainda mais longa, na figura 3.6). Por esse motivo, resolvemos trabalhar apenas com o modelo sem **S-S**.

A flexibilidade que a falta de **S-S** confere ao glóbulo, é apropriada para acessar vários níveis energéticos da proteína, o que não ocorre no modelo com **S-S** intactas, devido as restrições causadas pelas mesmas (observe a baixa variação em uma simulação ainda mais longa, na figura 3.15). Por esse motivo, resolvemos trabalhar apenas com o modelo sem **S-S**.

Esse período se mostrou suficiente para nos gerar dados os quais nos foram satisfatórios (Fig. 3.16). No caso do constructo sem pontes dissulfeto, o RMSD da proteína se mantém mais ou menos estável até os ~16000 ps, onde encontramos uma desestabilização nas folhas beta e parte da alfa hélice. Em ~22500 ps, a proteína chega ao primeiro limiar de sua resistência, desnaturando sua porção helical completamente até ~25000 ps. A partir disso, só se poderia esperar que a proteína estivesse inutilizada. No entanto, à partir de ~28000 ps, regiões helicais formam-se espontaneamente. Após essa breve renaturação, a proteína sofre a transposição de seu segundo limiar de resistência a ~46000 ps, sofrendo total desnaturação.

Figura 3.15. Deformação estrutural da Ts Kappa com **S-S** ao longo de 100ns. Aqui podemos perceber claramente como durante o tempo todo a proteína se manteve estável. Resultado provavelmente conseguido com a estabilização pelas das ligações covalentes **S-S**.

Figura 3.16. Em cima: Raio de giração da proteína. Em baixo: seu RMSD. As figuras acompanhando a trajetória são correspondentes visuais de como se encontram as estruturas referentes aos tempos na projeção do eixo das abscissas.

3.3.2 ReFolding em alta temperatura: perspectivas

Continuando a elongação do tempo de simulação, até 100 ns, obtivemos como resultado a (surpreendente) re-naturação (instantânea) em alta temperatura da Ts Kappa,

com a cadeia polipeptídica mostrando uma aproximação configuração da estrutura nativa muito significativa (Fig. 3.17, 3.18).

A partir de 57000 ps, a molécula começa a se aproximar cada vez mais de sua correspondente nativa, chegando em 80000 ps a um valor assintótico de RMSD $\sim 0,5$ nm, quase se igualando ao valor inicial de $\sim 0,4$ nm. Essa dinâmica intrincada é alvo primário de estudo no período a seguir, onde nos concentraremos nas análises da deformação de estruturas secundárias, com o objetivo de delimitação das forças determinantes ao direcionamento da molécula desnaturada ao estado *re-folded*.

Figura 3.17. RMSD para o Ts Kappa sem S-S ao longo de 100 ns, com temperatura de 398 K. Para discussão, por favor recorra ao texto.

Figura 3.18. Raio de giração para o constructo sem pontes dissulfeto ao longo de 100 ns, com temperatura de 398 K. Para discussão, recorra ao texto.

A figura 3.19 a seguir mostra *frames* instantâneos selecionados do filme completo de todo o processo de renaturação (de 50 a 100 ns), o qual consta no Anexo II. Note como ao cabo de aproximadamente 80ns a cadeia volta a assumir uma conformação muito próxima da nativa, embora por apenas uma parte da simulação.

Figura 3.19. Momentos da renaturação da Ts Kappa, a partir de 50ns até 100ns (Captura de tela). O vídeo completo se encontra no Anexo I. (verde: *coil*, vermelho: alfa hélice, amarelo: folha beta e roxo: 3,10 hélice) Detalhes maiores encontram-se no texto.

Para interpretarmos este resultado, retomamos a introdução deste Capítulo, quando apontamos para o fato experimental de que, no intervalo de temperatura entre 273-350K, aproximadamente, a variação da energia livre para se transferir um soluto

hidrofóbico do meio apolar para a água (ou ao contrário) é praticamente invariante, embora a entalpia e entropia variem consideravelmente [REF Lesk]. Assim, este resultado corrobora com a noção de que são as forças entrópicas (decorrentes do efeito hidrofóbico) as responsáveis para a indução da cadeia protéica à estrutura nativa. Assim, mesmo desnaturada e em alta temperatura, as forças entrópicas continuam atuando praticamente com a mesma intensidade, e a cadeia mais cedo ou mais tarde voltará a visitar as vizinhanças da estrutura nativa. E o processo continua ciclicamente, pois flutuações térmicas, a esta mais elevada temperatura, provocam a quebra das LH protegidas no interior da proteína, desfazendo a conformação globular da proteína.

3.4 Conclusão e comentários finais

Os resultados aqui apresentados são decorrentes da experimentação computacional envolvendo uma única proteína, a Ts-Kappa, que foi escolhida como um protótipo de proteína pequena (35 resíduos) e contendo três *cross-links* **S-S**. Portanto, somente considerando os tópicos aqui abarcados, assim como em praticamente todos os fenômenos envolvendo proteínas, dificilmente se pode generalizar. De fato, o acúmulo de evidências tomadas de observações independentes, sobre sistemas correlatos, é que possibilita a geração de hipóteses que deverão então se testadas rigorosamente.

Este foi o espírito deste trabalho: partiu-se de algumas premissas sobre o processo de *folding* e com a pretensão de se começar a desenvolver técnicas para a previsão teórica de estruturas espaciais de casos especiais de proteínas: aquelas proteínas pequenas e com restrições configuracionais, impostas pelas pontes dissulfeto (**S-S**). Para testar a hipótese do *folding* como um processo em duas etapas, a saber, a da busca pela estrutura nativa (governada essencialmente pelo efeito hidrofóbico), e as condições gerais de estabilidade (ditadas pelas pontes de hidrogênio), iniciou-se os trabalhos pela análise conjunta do papel das pontes **S-S** e das LH.

A existência de proteínas sem pontes **S-S** implica que somente LH podem estabilizar termicamente uma proteína, fato esse confirmado neste trabalho, com a simulação em várias temperaturas da Ts-Kappa destituída de suas 3 pontes **S-S**.

E por último, o fato de termos observado o *re-folding* da Ts-Kappa em solução a $T=398\text{K}$, numa janela de tempo de aproximadamente 100ns, corroboram as premissas de que o *folding* de proteínas ocorre por meio de um processo em duas etapas separadas no tempo: no início, forças entrópicas, decorrentes do efeito hidrofóbico,

dominam o cenário de busca pela estrutura nativa, e uma vez na vizinhança da nativa, as pontes de hidrogênio (protegidas da competição com o meio solvente), juntamente com uma eficiente especificidade estrutural das cadeias laterais (otimizando as forças de van der Waals), só aí iniciam a etapa de estabilização energética da proteína.

A possibilidade de re-*folding* térmico em temperaturas elevadas pode parecer um tanto artificial, uma vez que muitas proteínas se desnaturam exatamente pelo fato de mudanças nas condições fisiológicas, como a temperatura. Contudo, a ideia de que o *folding* se processa em etapas consecutivas, mas distinguidas pela atuação de diferentes composição de forças, pode eliminar esta percepção pois o processo de busca pela nativa quase não sofre alterações com a temperatura, enquanto que as condições de estabilidade pode ser muito mais dependente, mesmo porque uma proteína pode se desnaturada por uma variação negativa de temperatura (*cold denaturation*).

Porém, um argumento experimental nos é provido naturalmente a favor do *folding* a altas temperaturas. Trata-se da existência dos organismos termofílicos, os quais vivem e se reproduzem em temperaturas escaldantes. Nestas condições, suas proteínas encontram suas respectivas nativas e ficam nelas o tempo necessário. Inclusive este fato parece nos indicar que evolutivamente, levando em conta as condições primordiais na Terra, parece mesmo que as proteínas desenvolveram a capacidade de encontrar suas respectivas nativas em altas temperaturas, para depois evoluírem para reproduzirem suas propriedades em geral para condições ambientais mais parecidas com as atuais.

4.0 – Referências bibliográficas

- Allen, M.P.; Tildesley, D.J. Computer Simulation of Liquids. Clarendon Press, Oxford, 1987.
- Anon (2007). "Chaperones' influence on protein folding." Chemical & Engineering News **85**(49): 52-52.
- Alper, H. E. and R. M. Levy (1989). "Computer-Simulations of the Dielectric-Properties of Water - Studies of the Simple Point-Charge and Transferable Intermolecular Potential Models." Journal of Chemical Physics **91**(2): 1242-1251.
- Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(96): 223-30.
- Baez, L. A. and P. Clancy (1995). "Phase-Equilibria in Extended Simple Point-Charge Ice-Water Systems." Journal of Chemical Physics **103**(22): 9744-9755.
- Baker E. & Hubbard R. *Hydrogen bonding in globular proteins*. Prog. Biophy. Mol. Biol., **44**, 97-179 (1984).
- Bagchi, K., S. Balasubramanian, et al. (1997). "The effects of pressure on structural and dynamical properties of associated liquids: Molecular dynamics calculations for the extended simple point charge model of water." Journal of Chemical Physics **107**(20): 8561-8567.
- Barnett, R. N. and U. Landman (1993). "Born-Oppenheimer Molecular-Dynamics Simulations of Finite Systems - Structure and Dynamics of (H₂O)₂." Physical Review B **48**(4): 2081-2097.
- Bates, P. A., L. A. Kelley, et al. (2001). "Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM." Proteins-Structure Function and Genetics: 39-46.
- Beeman, D. (1976). "Some Multistep Methods for Use in Molecular-Dynamics Calculations." Journal of Computational Physics **20**(2): 130-139.
- Berendsen, H. J. C., J. P. M. Postma, et al. (1984). "Molecular-Dynamics with Coupling to an External Bath." Journal of Chemical Physics **81**(8): 3684-3690.
- Blanc, E., C. Lecomte, et al. (1997). "Solution structure of TsKapa, a charybdotoxin-like scorpion toxin from Tityus serrulatus with high affinity for apamin-sensitive Ca²⁺-activated K⁺ channels." Proteins-Structure Function and Genetics **29**(3): 359-369.
- Blattner, F. R., G. Plunkett, et al. (1997). "The complete genome sequence of Escherichia coli K-12." Science **277**(5331): 1453-&.

- Browne, W. J.; North, A.C.; Phillips, D.C.; Brew, K.; Vanaman, T.C.; Hill, R. L. A (1969) Possible three-dimensional structure of bovine {alpha}-lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.* **42**: 65–86.
- Bryngelson, J. D., J. N. Onuchic, et al. (1995). "Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis." Proteins-Structure Function and Genetics **21**(3): 167-195.
- Chang, I., M. Cieplak, et al. (2001). "Protein threading by learning." Proceedings of the National Academy of Sciences of the United States of America **98**(25): 14350-14355.
- Chaudhuri, T. K. and S. Paul (2006). "Protein-misfolding diseases and chaperone-based therapeutic approaches." FEBS J **273**(7): 1331-49.
- Chen, B., M. G. Martin, et al. (1998). "Thermodynamic properties of the Williams, OPLS-AA, and MMFF94 all-atom force fields for normal alkanes." Journal of Physical Chemistry B **102**(14): 2578-2586.
- Claessens, M., E. Vancutsem, et al. (1989). "Modeling the Polypeptide Backbone with Spare Parts from Known Protein Structures." Protein Engineering **2**(5): 335-345.
- Crozier, P. S., D. Henclerson, et al. (2001). "Model channel ion currents in NaCl-extended simple point charge water solution with applied-field molecular dynamics." Biophysical Journal **81**(6): 3077-3089.
- Dahiyat, B. I. and S. L. Mayo (1997). "De novo protein design: Fully automated sequence selection." Science **278**(5335): 82-87.
- Das, P., S. Matysiak, et al. (2005). "Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes." Proceedings of the National Academy of Sciences of the United States of America **102**(29): 10141-10146.
- Eriksson, A., M. N. Jacobi, et al. (2008). "Effective thermostat induced by coarse graining of simple point charge water." Journal of Chemical Physics **129**(2): -.
- Fasman, G. D. (ED.). 1989. Prediction of Protein Structure and the Principle of Protein Conformation. *Plenum Press*. New York.
- Fersht, A., Structure and Mechanism in Protein Science: A Guide to Enzyme. Catalysis and Protein Folding, W.H. Freeman, New York. 1999.
- Floudas, C. A., H. K. Fung, et al. (2006). "Advances in protein structure prediction and de novo protein design: A review." Chemical Engineering Science **61**(3): 966-988.
- Gellman, S. H. (1998). "Foldamers: A manifesto." Accounts of Chemical Research **31**(4): 173-180.

- Gennes, P. G. de (1985). "Scaling Concepts in Polymer Physics." Cornell University Press, Ithaca, New York.
- Ghelis, C.; Yon, J. (1982). Protein folding. *Academic Press*. New York.
- Hartl, F. U. (1996). "Molecular chaperones in cellular protein folding." Nature **381**(6583): 571-580.
- Hartl, F. U. and J. Martin (1995). "Molecular Chaperones in Cellular Protein-Folding." Current Opinion in Structural Biology **5**(1): 92-102.
- Henikoff, S. and J. G. Henikoff (1992). "Amino-Acid Substitution Matrices from Protein Blocks." Proceedings of the National Academy of Sciences of the United States of America **89**(22): 10915-10919.
- Hess, B., C. Kutzner, et al. (2008). "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation." Journal of Chemical Theory and Computation **4**(3): 435-447.
- Hesterkamp, T. and B. Bukau (1998). "Role of the DnaK and HscA homologs of Hsp70 chaperones in protein folding in E-coli." Embo Journal **17**(16): 4818-4828.
- Hill, D. J., M. J. Mio, et al. (2001). "A field guide to foldamers." Chemical Reviews **101**(12): 3893-4011.
- Ho, B. K. and K. A. Dill (2006). "Folding very short peptides using molecular dynamics." Plos Computational Biology **2**(4): 228-237.
- Ho, B. K. and K. A. Dill (2006). "Folding very short peptides using molecular dynamics (vol 2, pg 5, 2006)." Plos Computational Biology **2**(5): 476-477.
- Hockney, R. W. (1970). *Methods in Computational Physics* vol. **9**, Academic Press, New York pp. 135–211
- Imberty, A., J. Gruza, et al. (1998). "The crystal and molecular structure of a diglycosylamine: the N-analogue of peracetylated alpha,beta-trehalose." Carbohydrate Research **311**(3): 135-146.
- Irback, A. and S. Mohanty (2006). "PROFASI: A Monte Carlo simulation package for protein folding and aggregation." Journal of Computational Chemistry **27**(13): 1548-1555.
- Jorgensen, W. L. and J. Tirado-Rives (1998). "Development of the OPLS-AA force field for organic and biomolecular systems." Abstracts of Papers of the American Chemical Society **216**: U696-U696.
- Kahn, K. and T. C. Bruice (2002). "Parameterization of OPLS-AA force field for the conformational analysis of macrocyclic polyketides." Journal of Computational Chemistry **23**(10): 977-996.

- Kaminski, G. and W. L. Jorgensen (1996). "Performance of the AMBER94, MMFF94, and OPLS-AA force fields for modeling organic liquids." Journal of Physical Chemistry **100**(46): 18010-18013.
- Karasawa, T., K. Tabuchi, et al. (1993). "Development of Simulation-Models for Protein-Folding in a Thermal Annealing Process .1. A Simulation of Bpti Folding by the Pearl Necklace Model." Computer Applications in the Biosciences **9**(3): 243-251.
- Koehl, P. and M. Delarue (1996). "Mean-field minimization methods for biological macromolecules." Current Opinion in Structural Biology **6**(2): 222-226.
- Kolinski, A., W. Galazka, et al. (1996). "On the origin of the cooperativity of protein folding: Implications from model simulations." Proteins-Structure Function and Genetics **26**(3): 271-287.
- Kolinski, A., J. Skolnick, et al. (1996). "A method for the prediction of surface loops/turns and transglobular connections in small proteins." Abstracts of Papers of the American Chemical Society **211**: 34-Comp.
- Kony, D., W. Damm, et al. (2002). "An improved OPLS-AA force field for carbohydrates." Journal of Computational Chemistry **23**(15): 1416-1429.
- Kubach, C. and N. Rougeau (1998). "On the use of a Born-Oppenheimer type separation in the treatment of the dynamics of molecular collisions." Journal of Molecular Structure-Theochem **424**(1-2): 171-180.
- Leach, A. (2001). Molecular Modelling : Principles and Applications. 2ª Edição. Prentice Hall.
- Lee, B. (1991). "Solvent Reorganization Contribution to the Transfer Thermodynamics of Small Nonpolar Molecules." Biopolymers **31**(8): 993-1008.
- Leopold, P. E., M. Montal, et al. (1992). "Protein Folding Funnels - a Kinetic Approach to the Sequence Structure Relationship." Proceedings of the National Academy of Sciences of the United States of America **89**(18): 8721-8725.
- Lesk, A. M. (2003). "Hydrophobicity - getting into hot water." Biophysical Chemistry **105**(2-3): 179-182.
- Levinthal, C. (1968). "Are There Pathways for Protein Folding." Journal De Chimie Physique Et De Physico-Chimie Biologique **65**(1): 44-&.
- Liwo, A., M. Khalili, et al. (2005). "Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains." Proceedings of the National Academy of Sciences of the United States of America **102**(7): 2362-2367.

- Lund, O.; Nielsen, M.; Lundegaard, C.; Worning, P. (2002) CPHmodels 2.0: X3M a Computer Program to Extract 3D Models. Resumo da CASP5 Conference A102.
- Martin, J. (1997). "Molecular chaperones and mitochondrial protein folding." Journal of Bioenergetics and Biomembranes **29**(1): 35-43.
- Martin, M. G. (2006). "Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor-liquid coexistence curves and liquid densities." Fluid Phase Equilibria **248**(1): 50-55.
- Martin, M. G. and J. I. Siepmann (1998). "Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes." Journal of Physical Chemistry B **102**(14): 2569-2577.
- Nakamura, T. and S. A. Lipton (2009). "Cell death: protein misfolding and neurodegenerative diseases." Apoptosis **14**(4): 455-68.
- Nelson, J. C., J. G. Saven, et al. (1997). "Solvophobically driven folding of nonbiological oligomers." Science **277**(5333): 1793-1796.
- Ogata, K. and H. Umeyama (2000). "An automatic homology modeling method consisting of database searches and simulated annealing." Journal of Molecular Graphics & Modelling **18**(3): 258-+.
- Oh, K., K. S. Jeong, et al. (2001). "Folding-driven synthesis of oligomers." Nature **414**(6866): 889-893.
- Onuchic, J. N., Z. LutheySchulten, et al. (1997). "Theory of protein folding: The energy landscape perspective." Annual Review of Physical Chemistry **48**: 545-600.
- Onuchic, J. N., N. D. Socci, et al. (1996). "Protein folding funnels: The nature of the transition state ensemble." Folding & Design **1**(6): 441-450.
- Pauling, L. (1960). Chemical Bond. In: PAULING, L. (Ed), The Nature of Chemical Bond, 3rd Edn. New York: Cornell University Press. p. 85
- Pennisi, E. (1997). "Organic chemistry - Polymer folds just like a protein." Science **277**(5333): 1764-1764.
- Petrey, D., Z. X. Xiang, et al. (2003). "Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling." Proteins-Structure Function and Genetics **53**(6): 430-435.
- Ptitsyn, O. B., A. V. Finkelstein, et al. (1999). "Protein folding: A bridge between physics and biology." Molecular Biology **33**(6): 893-896.
- Ramos, C. H. and S. T. Ferreira (2005). "Protein folding, misfolding and aggregation: evolving concepts and conformational diseases." Protein Pept Lett **12**(3): 213-22.

- Rhee, H. S. and C. H. Yang (2005). "A protein engineering study of the role of glutamine 19 in the catalytic mechanism of actinidin." Febs Journal **272**: 18-18.
- Sali, A. and T. L. Blundell (1993). "Comparative Protein Modeling by Satisfaction of Spatial Restraints." Journal of Molecular Biology **234**(3): 779-815.
- Santiveri, C. M., E. Leon, et al. (2008). "Context-dependence of the contribution of disulfide bonds to beta-hairpin stability." Chemistry-a European Journal **14**(2): 488-499.
- Scheraga, H. A., M. Khalili, et al. (2007). "Protein-folding dynamics: Overview of molecular simulation techniques." Annual Review of Physical Chemistry **58**: 57-83.
- Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: an automated protein homology-modeling server." Nucleic Acids Research **31**(13): 3381-3385.
- Scott, W. R. P., P. H. Hunenberger, et al. (1999). "The GROMOS biomolecular simulation program package." Journal of Physical Chemistry A **103**(19): 3596-3607.
- Seshasayee, A. S. N. (2005) High-Temperature unfolding of a trp-Cage mini-protein: a molecular dynamics simulation study. Theor Biol Med Model. **2**: 7.
- Siew, N., A. Elofsson, et al. (2000). "MaxSub: an automated measure for the assessment of protein structure prediction quality." Bioinformatics **16**(9): 776-785.
- Sparrer, H., K. Rutkat, et al. (1997). "Catalysis of protein folding by symmetric chaperone complexes." Proceedings of the National Academy of Sciences of the United States of America **94**(4): 1096-1100.
- van der Spoel, D; Lindahl, E., Hess, B. Gromacs user manual V. 3.3. in <http://wiki.gromacs.org>, 2006.
- Suardiaz, R., J. Perez, et al. (2006). "A DFT study of the substituent induced shifts of a 5 alpha-OH on ring A C-13 chemical shifts of spirostane sapogenins." Journal of Molecular Structure-Theochem **769**(1-3): 87-89.
- Swope, W. C., H. C. Andersen, et al. (1982). "A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters." Journal of Chemical Physics **76**(1): 637-649.
- Torda, A. E. (2004). "Protein sequence optimization - Theory, practice, and fundamental impossibility." Soft Materials **2**(1): 1-10.
- Torda, A. E., J. B. Procter, et al. (2004). "Wurst: a protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices." Nucleic Acids Research **32**: W532-W535.

- Ulmschneider, J. P., M. B. Ulmschneider, et al. (2006). "Monte Carlo vs molecular dynamics for all-atom polypeptide folding simulations." Journal of Physical Chemistry B **110**(33): 16733-16742.
- Verlet, L. (1967). "Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules." Physical Review **159**(1): 98-&.
- Villamanan, R. M., C. R. Chamorro, et al. (2006). "Phase equilibria properties of binary and ternary systems containing di-isopropyl ether plus isobutanol plus benzene at 313.15 K." Fluid Phase Equilibria **239**(2): 178-182.
- Wallner, B. and A. Elofsson (2005). "All are not equal: A benchmark of different homology modeling programs." Protein Science **14**(5): 1315-1327.
- Welch, W. J. and C. R. Brown (1996). "Influence of molecular and chemical chaperones on protein folding (vol 1, pg 109, 1996)." Cell Stress & Chaperones **1**(3): 207-207.
- Wensink, E. J. W., A. C. Hoffmann, et al. (2003). "Dynamic properties of water/alcohol mixtures studied by computer simulation." Journal of Chemical Physics **119**(14): 7308-7317.
- Xu, Z. T., H. H. Luo, et al. (2007). "Modifying the OPLS-AA force field to improve hydration free energies for several amino acid side chains using new atomic charges and an off-plane charge model for aromatic residues." Journal of Computational Chemistry **28**(3): 689-697.
- Yooseph, S., G. Sutton, et al. (2007). "The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families." Plos Biology **5**(3): 432-466.
- Zhou, H., W. Wang, et al. (2005). "Contributions of disulfide bonds in a nested pattern to the structure, stability, and biological functions of endostatin." Journal of Biological Chemistry **280**(12): 11303-11312.