

UNIVERSIDADE DE SÃO PAULO
Faculdade de Filosofia, Ciências e Letras de Ribeirão
Preto, Departamento de Física e Matemática

SIMULAÇÕES MICROCANÔNICAS DE PROTEÍNAS

Rafael Bertolini Frigori

Tese apresentada à Faculdade
de Filosofia, Ciências e Letras
de Ribeirão Preto da USP,
como parte das exigências para
a obtenção do título de Doutor
em Ciências: Física Aplicada à
Medicina e Biologia.

Orientador:

Prof. Dr. Nelson A. Alves

Ribeirão Preto, dezembro de 2010



”Wenn Sie mich nach meiner innersten Überzeugung fragen ob man unser (das 19.) Jahrhundert einmal das eiserne Jahrhundert oder das Jahrhundert des Dampfes oder der Elektrizität nennen wird, so antworte a ich ohne Bedenken, das Jahrhundert der mechanischen Naturauffassung, das Jahrhundert Darwins wird es heißen”
 [L. BOLTZMANN, Populäre Schriften, 1886]¹

1

Se você perguntar sobre minha mais profunda convicção se nosso século (XIX) será chamado de o século do ferro ou o século do vapor ou o da eletricidade, responderei sem hesitação: ele será chamado o século da visão mecânica da natureza, o século de Darwin!

Agradecimentos

A todos os colegas, amigos e familiares que de algum modo colaboraram para o sucesso dessa minha nova empreitada. E em particular:

A *Deus* pela gênese deste "pequeno" Universo que tanto nos fascina investigar;

Ao Prof. Nelson, não só pelos bons exemplos, paciência e dedicação à orientação deste doutorado mas também pela amizade e compreensão nos momentos mais difíceis;

Ao amigo Leandro G. Rizzi pelo companherismo e valorosa parceria científica;

Aos funcionários e professores do DFM/FFCLRP/USP por cultivarem um ambiente acolhedor e propício à pesquisa científica;

Aos "velhos" amigos Cesar A.V. Moura, Fábio V. Boas, Felipe L. Scanavini, Guilherme C.P. Innocentini e Tiago M. Franco; por vezes distantes, mas sempre presentes;

Aos colegas e amigos do departamento: André S.C. Peres, Aquino L. Espindola, Ariadne de A. Costa, Brenno T. Cabella, César A.S. Terçariol, Denise de Arruda, Diogo Porfírio, Ebenézer S. Cavalcanti, Fabiano L. Ribeiro, Fernanda M. Oliveira, Guidolins (Leila B.M. & Luis C.B.M.-), Jacyana S.M. Fonseca, Juliana M.S. Berbert, Lindomar S. Santos, Luis A. Cabral, Marcelo A. Pereira, Matheus R. Mendonça, Matheus S. de Moura, Natália Destefano, Olavo H. Menin, Raimundo N.A. Costa, Rodrigo S. Gonzalez, Tiago J. Arruda e Wilnice T.R. Oliveira pelos bons momentos compartilhados;

À Ana Bonfleur pelo carinho e apoio durante a conclusão desta tese;

Ao pessoal da UTFPR-Toledo pela acolhida na minha nova "casa";

Ao povo brasileiro, que por meio da CAPES, financiou a execução deste trabalho;

À Carla B.F. Junqueira e ao Olavo G. Junqueira pela amizade e incentivo;

Aos meus pais, Claudia R.B. Frigori e Emilio A.O. Frigori, pelo amor incondicional.

... I recognize that many physicists are smarter than I am — most of them theoretical physicists. A lot of smart people have gone into theoretical physics, therefore the field is extremely competitive. I console myself with the thought that although they may be smarter and may be deeper thinkers than I am, I have broader interests than they have ...

[LINUS PAULING, *The Meaning of Life*, 1990]

À memória de Marco Alberto Perez[†]
físico, professor e saudoso amigo

Resumo

Transições de fase termodinâmicas são usualmente estudadas por meio do ensemble canônico e estão associadas a sistemas macroscópicos. Entretanto, tem-se tornado cada vez mais frequente e importante o estudo de sistemas físicos "pequenos", cujos alcances característicos das interações equivalem aos tamanhos dos sistemas. Nestes casos pode haver inequivalência entre grandezas físicas obtidas nos ensembles canônico e microcanônico. Encontramos inúmeros exemplos destes sistemas em diversas áreas da Física. Na área de matéria condensada temos, por exemplo, o modelo Blume-Capel com interações de alcance infinito. Neste modelo as soluções exatas canônica e microcanônica são inequivalentes. Este modelo é investigado nesta tese por meio de um ensemble interpolante, conhecido como gaussiano estendido, como uma aplicação teórica preliminar. Adicionalmente, empregamos o arcabouço mecânico-estatístico no estudo de certas biomoléculas com ampla importância biológica: proteínas. Atualmente o estudo do comportamento termodinâmico destas moléculas tem ficado restrito quase que exclusivamente a abordagem via ensemble canônico. Neste trabalho analisamos os resultados da etapa microcanônica de simulações multicanônicas procurando obter aspectos físicos de biomoléculas como os domínios Src SH3 (pdb: 1NLO) e as Príons humanas (pdb: 1HJM). Caracterizamos com esta abordagem as transições de fase de enovelamento e de agregação destes sistemas. Os resultados obtidos são interpretados à luz da termoestatística microcanônica, oferecendo um ponto de vista fenomenológico alternativo à abordagem usual.

Abstract

Thermodynamic phase transitions are usually studied by the canonical ensemble and they are associated to macroscopic systems. However, it is becoming more frequent and important the study of "small" physical systems: whose characteristic interaction-lengths are equivalent to system sizes. In these cases there can happen inequivalences among quantities computed in the canonical and microcanonical ensembles. There are innumerable examples of that systems in various areas of physics. In the field of condensed matter there is for instance the Blume-Capel model with infinite-range interactions. The canonical and microcanonical exact solutions of this model are inequivalents. That model is investigated on this thesis through an interpolating ensemble, known as the extended gaussian, as a preliminar theoretical application. Additionally, we have employed the statistic-mechanical framework to study some biomolecules of large biological interest: proteins. Nowadays the study of the thermodynamic behavior of that molecules has been restricted almost only to the canonical approach. However, in this work we have analysed by the microcanonical step of multicanonical simulations the physical aspects of biomolecules as the domain Src SH3 (pdb: 1NLO) and the human Prions (pdb: 1HJM). Thus, we characterize with this approach the phase transitions of folding and aggregation of that systems. The results obtained are interpreted under the light of the microcanonical thermostatics, offering an alternative phenomenological viewpoint.

Lista de Figuras

2.1	Esquerda: situação que exemplifica uma entropia côncava, i.e., fase única. Direita: ilustração de um intruso convexo na entropia, como o que ocorre nas transições canônicas de primeira ordem; aqui os dois pontos-de-sela correspondem a duas fases distintas. Adaptado da referência [2].	20
4.1	Estrutura geral de um aminoácido, com o carbono alfa C_α destacado ao centro, o grupo amina (NH_2) à esquerda, o carboxila (CO_2H) à direita e a cadeia lateral R abaixo.	36
4.2	Estrutura química, nomenclatura, abreviação com código de 3 letras e classificação quanto a hidrofobicidade dos 20 aminoácidos primários.	37
4.3	Formação de um dímero (dipeptídeo) pela condensação de aminoácidos.	38
4.4	Hierarquia de estruturas em proteínas. (a) Primária: com abreviações de uma letra. (b) Secundária: (1) hélice alfa e (2) fita beta. (c) Terciária: associação de motivos enovelados, (d) Quaternária: união de cadeias peptídicas monoméricas.	39
4.5	Estrutura secundária: arranjo espacial típico de uma <i>hélice alfa</i> , note as cadeias laterais externas ao eixo de simetria e a frequência espacial de 3,6 resíduos por volta [23].	40
4.6	Estrutura secundária: conformações usuais de <i>folhas beta</i> , notam-se as fitas orientadas antiparalelas (superior) e paralelas (inferior), além das terminações amina (N) e carboxila (C).	41
4.7	O perfil da energia livre (<i>energy landscape</i>) em formato de funil é um modelo aceito para a descrição do enovelamento protéico. Os estados 11, 12, 13 são intermediários enquanto "F" é a estrutura protéica nativa, ou de menor energia livre. As rotas preferenciais são representadas por setas.	43
4.8	Proteínas envolvidas na formação de agregados. Superior à esquerda: <i>Alfa-Sinucleína</i> , associada ao mal de Parkinson. Superior à direita: <i>Amilina</i> , cuja formação de agregados ocorre na Diabetes do Tipo II. Inferior à esquerda: proteína precursora do <i>Amiloide-Beta</i> ($A\beta$), agregados de $A\beta$ ocasionam o mal de Alzheimer. Inferior à direita: domínio Src SH3, associado à transdução de sinais, não patológica. Adaptado do protein data bank (PDB).	45
4.9	Modelo de fibrilas amilóides, ricas em formações de fitas beta como as hélices beta-cruzadas, que possuem alta resistência à desnaturação, [33].	46

4.10	Representação esquemática dos possíveis estados conformacionais assumidos por cadeias peptídicas e suas interconversões usuais. Adaptado da Ref. [33].	47
4.11	Configurações assumidas por Prions: a forma usual PrP^C representada por (A); e a patogênica PrP^{sc} associada aos males de Creutzfeld-Jacob e à Encefalopatia Espongiforme ilustrada em (B).	48
4.12	Esquema de um único peptídeo no modelo AB com N resíduos [46, 65, 87]. Note a representação da cadeia principal dos carbonos alfa (esferas), assim como seus $2N$ graus de liberdade de rotação (i.e. ângulos θ e φ).	51
5.1	Comportamento no limite microcanônico da solução do modelo BC via formalismo EGE, com acoplamento $\Delta/J = 0.462407$, correspondendo à região de transição de fase de primeira ordem canônica e de segunda ordem microcanônica. (a) Temperatura microcanônica como função da energia média ε . A linha tracejada horizontal corresponde à temperatura crítica canônica. (b) A entropia deslocada $\tilde{s}(\varepsilon) = s_{micro}(\varepsilon) - (A + B\varepsilon)$, com $A = 0.401447$ e $B = 1.398397$. A subtração é efetuada para uma melhor visualização da não concavidade da entropia em relação a função linear ligando $s(\varepsilon_a)$ a $s(\varepsilon_b)$, com $\varepsilon_a = 0.328959$ e $\varepsilon_b = 0.330646$. (c) Calor específico $c(\varepsilon)$. Ele apresenta dois polos localizados pelos zeros do determinante $d_{S_{micro}}(\varepsilon, m)$, onde m denota os valores da magnetização que maximizam a entropia em dado ε . Esses pólos também podem ser observados a partir do comportamento de $T(\varepsilon)$ em (a). (d) Susceptibilidade magnética $\chi(\varepsilon)$. Ela apresenta dois polos, novamente posicionados nos zeros do determinante $d_{S_{micro}}(\varepsilon, m)$ e torna-se negativa entre eles. (e) Comportamento do determinante $d_{S_{micro}}(\varepsilon, m)$ como função de ε . As linhas tracejadas verticais mostram os zeros de $d_{S_{micro}}(\varepsilon, m)$	58
5.2	Entropia $s(\varepsilon, m)$ para alguns valores de ε e m com $\Delta/J = 0.462407$. Para valores inferiores da energia ε , alguns intervalos da magnetização são inacessíveis; o que demonstra quebra de ergodicidade neste modelo. Domínios magnéticos desconexos parecem ser típicos de sistemas com interações de longo alcance, exibindo transições de fase (canônica) de primeira ordem.	59
5.3	Curva calórica $T(\varepsilon) \times \varepsilon$ computadas para diversos valores de γ a partir da solução geral do modelo Blume-Capel no EGE. O acoplamento empregado é $\Delta/J = 0.462407$ e no limite de $\gamma \rightarrow \infty$ reobtemos a curva microcanônica (a) da Figura (5.1).	60
5.4	Valores mínimos que o parâmetro $\tilde{\gamma}$, calculado segundo a Eq. (5.19), deve assumir para que a termodinâmica deduzida via EGE para o modelo Blume-Capel (com $\Delta/J = 0.462407$) seja equivalente à microcanônica.	60

- 5.5 (A) Temperaturas EGE obtidas no limite canônico ($\gamma = 0$) para alguns valores de Δ/J . Para $\Delta/J = 0.462098$ temos simultaneamente transições de fase canônica e microcanônica de segunda ordem. Quando $\Delta/J = 0.4622$ e $\Delta/J = 0.4623$ as transições de fase canônicas são de primeira ordem, mas as microcanônicas são ainda de segunda ordem. (B) Temperaturas EGE no limite canônico ($\gamma = 0$), todos os valores de Δ/J estão na região de transição de primeira ordem canônica e microcanônica. (C) Temperaturas computadas via EGE no limite microcanônico ($\gamma \cong 10^{12}$), para $\Delta/J = 0.4622$ e $\Delta/J = 0.4623$ em que ocorrem transições de fase microcanônicas de segunda ordem e canônicas de primeira ordem. (D) Temperaturas EGE no limite microcanônico ($\gamma \cong 10^{12}$) na região de acoplamentos $\Delta/J = 0.4625$ e $\Delta/J = 0.4627$, cujas transição de fase são de primeira ordem. 61
- 5.6 Na figura a esquerda temos linhas de transição de fase canônicas. A linha crítica (cinza pontilhada) termina no ponto tricrítico canônico \bullet , a partir do qual a transição torna-se de primeira ordem (linha cheia). A figura menor é uma ampliação que mostra a linha (cheia) da transição de fase de primeira ordem canônica e as linhas (ponto-tracejadas) da transição de fase de primeira ordem microcanônica. Na figura a direita temos uma representação esquemática do diagrama de fase do modelo Blume-Capel, ampliado ao redor dos Pontos Tricrítico Canônico (CTP) e Microcanônico (MTP). A linha de transição de fase de segunda ordem (comum a ambos ensembles) é pontilhada; a linha canônica de primeira ordem é cheia e as linhas microcanônicas tracejadas são de segunda ordem em negrito e de primeira ordem em cinza. Adaptado da referência [10]. 63
- 5.7 Valor mínimo do parâmetro interpolante γ para que a solução EGE seja termodinamicamente equivalente à microcanônica entre os pontos tricríticos canônico ($\Delta/J \cong 0.4621$) e microcanônico ($\Delta/J \cong 0.4624$). 63
- 6.1 Esquema da atualização de Monte Carlo empregada em nossos estudos. Uma proteína posicionada a distância $|\vec{R}|$ da origem com ângulo polar "a" e equatorial "b" é atualizada para outra posição $\vec{R} \rightarrow \vec{R}' : \{|\vec{R}'|, a', b'\}$. Por outro lado, simultaneamente a posição relativa de cada um dos $(N + 1)$ -ésimos aminoácidos, em relação aos N -ésimos aminoácidos, é alterada de modo que $\vec{r} \rightarrow \vec{r}' = \vec{r} + \vec{d}r$. 65
- 6.2 Estrutura nativa de *backbone* de sequências Fibonacci. (Painel esquerdo) configuração com energia mínima, $E_{1 \times Fib.} = -5,75$ e $\Delta E = 0,1$ para a sequência artificial com 13 resíduos $Fibonacci_{13} : "ABBABBABBBAB"$. (Painel direito) configuração de energia mínima para a estrutura agregada $E_{2 \times Fib.} = -29,15$ e $\Delta E = 0,1$ de duas sequências idênticas de $Fibonacci_{13/AB}$ 67

- 6.3 Termodinâmica microcanônica para uma única sequência Fibonacci. (Painel superior): curva calórica $\beta \times E$, barras de erro correspondem ao desvio padrão de 50 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas pela aplicação recursiva de k filtragens do tipo média móvel de 10 pontos. (Painel central): cálculo da derivada da curva interpolante (vermelha) de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se que ocorre um pico positivo caracterizando uma transição de fase contínua, ou de segunda ordem, associada ao "enovelamento". 68
- 6.4 Termodinâmica microcanônica para duas sequências Fibonacci interagentes. (Painel superior): curva calórica $\beta(E) \times E$, barras de erro correspondem ao desvio padrão de 50 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas pela aplicação recursiva de k filtragens do tipo média móvel de 10 pontos. (Painel central): cálculo da derivada da curva interpolante (vermelha) de $\beta(E) \times E$, ou seja $\frac{d\beta(E)}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se uma transição de fase de primeira ordem (de agregação), com calores específicos negativos. Os pequenos picos positivos assinalam "formação de domínios" ou "enovelamento". 69
- 6.5 Estrutura nativa de *backbone* para domínios Src SH3 com 56 resíduos e código 1NLO no Protein Data Bank (PDB). Foi mapeada no modelo AB segundo sua natureza hidrofóbica/polar, resultando na sequência $SH3_{56/AB}$: "BAABABBB-BAABBBBAAABBABAABAABBBBABAAABBAABBABABABABBABBA". (Painel esquerdo): configuração nativa, ou com energia mínima $E_{1 \times SH3} = -41,48$ e $\Delta E = 1,0$ para a uma única sequência Src SH3. (Painel direito): configuração de energia mínima com $E_{2 \times SH3} = -29,15$ e $\Delta E = 1,0$ da forma agregada de dois peptídeos $SH3_{56/AB}$ interagentes. 72
- 6.6 Termodinâmica microcanônica para um único domínio Src SH3. (Painel superior): curva calórica $\beta \times E$, barras de erro correspondem ao desvio padrão de 54 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas por aplicações recursivas de filtros de média móvel de 5 pontos. (Painel central): cálculo da derivada de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se que ocorrem dois picos positivos indicando transições de fase contínuas (ou de segunda ordem). Biologicamente assinalariam o enovelamento (pico menor) e formação de estruturas secundárias (pico maior). 73
- 6.7 Termodinâmica microcanônica para dois domínios Src SH3 interagentes. (Painel Superior): curva calórica $\beta \times E$, barras vem do desvio padrão de 26 conjuntos de parâmetros MUCA. Linhas azul/vermelho emergem de 1 a 7 filtragens recursivas de médias móveis (de 5 pontos). (Painel Central): é a derivada da curva interpolante (vermelha) de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel Inferior): calor específico microcanônico do sistema ("filtro 7x5"). Há uma transição de primeira ordem (canônica, de agregação), com calores específicos microcanônicos negativos, e uma de segunda ordem (de enovelamento). 74

6.8	<p>Estrutura de <i>backbone</i> de príons humanas com 104 resíduos e código 1HJM no Protein Data Bank (PDB). Foi mapeada no modelo AB segundo sua natureza hidrofóbica/polar, resultando na sequência $Prion_{104/AB}$: "AAABAAAABAAAABAABAAA BBBBABBABBABBBBBABBABABBBABBBBBAAABBAABAB AB BBBBABB BBBBABB BBBBABAABAAB- BAAABBBBABAB". (Painel esquerdo): configuração nativa, ou com energia mínima, $E = -8,35$ e $\Delta E = 1,0$ para a sequência $Prion_{104/AB}$. (Painel direito): configuração de energia mínima, dita na forma agregada ou dimerizada, com $E = 12,59$ e $\Delta E = 1,0$ para dois peptídeos $Prion_{104/AB}$ interagentes.</p>	76
6.9	<p>Termodinâmica microcanônica para uma proteína 1HJM, a Príon humana. (Painel superior): curva calórica $\beta \times E$, barras de erro correspondem ao desvio padrão de 22 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas por filtragens recursivas do tipo média móvel de 5 pontos. (Painel central): cálculo da derivada de curvas interpolantes de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se que ocorre um único pico positivo (estável), caracterizando uma transição de fase contínua (ou de segunda ordem), associada ao "enovelamento sem formação de intermediários": um fenômeno já observado <i>in vitro</i> [73, 74].</p>	77
6.10	<p>Termodinâmica microcanônica para duas proteínas 1HJM, ou Príons humanas. (Painel superior): curva calórica $\beta \times E$, barras de erro são o desvio padrão de 10 conjuntos de parâmetros MUCA. Linhas em verde/vermelho emergem de 1 a 10 filtragens recursivas de médias móveis (de 5 pontos). (Painel central): é a derivada da curva interpolante (vermelha) de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema ("filtro 10x5"). Há apenas uma transição de primeira ordem (canônica, de agregação), com calores específicos negativos. Nenhuma outra transição de fase (e.g. contínua) foi observada, apesar de grandes esforços computacionais.</p>	78
7.1	<p>O gráfico acima mostra um conjunto de dados (100 pontos, em preto) e suas médias móveis com tamanhos diferentes [5]. Temos respectivamente as seguintes curvas: 2-pontos (vermelha), 4-pontos (amarela), 6-pontos (verde) e 8-pontos (azul). Note o efeito de filtragem do ruído nestas curvas.</p>	85
7.2	<p>Esquema termodinâmico utilizado na dedução do ensemble canônico. O sistema 1 está acoplado energeticamente a um reservatório térmico, dito sistema 2, ambos estão à mesma temperatura de equilíbrio T e são isolados do resto do Universo.</p>	87

Conteúdo

Lista de Figuras	7
Conteúdo	12
1 Introdução	14
2 Termoestatística microcanônica	17
2.1 O ensemble microcanônico	18
2.2 Transições de fase microcanônicas	19
2.3 Inequivalência de ensembles e transições de fase	21
2.4 Ensembles generalizados	23
2.4.1 O ensemble multicanônico	23
2.4.2 O ensemble gaussiano estendido	25
3 Simulações	28
3.1 Métodos de Monte Carlo markovianos	29
3.1.1 O algoritmo de Metropolis	30
3.1.2 Erros numéricos	31
3.2 Simulações microcanônicas	32
3.2.1 Relações de recorrência	33
3.2.2 Implementação alternativa	34
4 Proteínas e o modelo AB	35
4.1 Proteínas <i>in vitro</i> e <i>in vivo</i> : um panorama	38
4.1.1 Estrutura	39
4.1.2 Funcionalidades	42
4.1.3 Enovelamento protéico	42
4.2 Agregação protéica e proteinopatias	44
4.3 Proteínas <i>in silico</i>	49
4.3.1 O modelo AB <i>coarse-grained</i>	50
5 Resultados exatos: modelo de spin	52
5.1 O modelo Blume-Capel no ensemble gaussiano estendido	52
5.1.1 Um exemplo de sistemas "pequenos"	53

<i>Conteúdo</i>	13
5.2 Solução exata no ensemble gaussiano estendido	53
5.3 Limites termodinâmicos: inequivalência de ensembles	55
5.4 Pontos tricríticos	62
6 Resultados numéricos: proteínas	64
6.1 Implementação computacional	65
6.2 Enovelamento e agregação	66
6.2.1 Sequências Fibonacci	66
6.2.2 Domínios Src SH3	70
6.2.3 Príons humanas	72
7 Considerações finais	79
Apêndices	82
Bibliografia	89

Capítulo 1

Introdução

... What an organism feeds upon is negative entropy. Or to put it less paradoxically, the essential thing in metabolism is that the organism succeeds in freeing itself from all the entropy it cannot help producing while alive ...

[E. SCHRÖDINGER, What is life?, 1944]

A mecânica estatística microcanônica, como formulada por Ludwig Boltzmann [1, 2], constitui um dos principais pilares da moderna abordagem física para sistemas de muitos corpos. Originalmente concebida com o intuito de alcançar uma explicação microscópica, ou cinética, da termodinâmica dos gases, ela permitiu inferir a existência de átomos décadas antes da sua observação experimental. Aqui, a entropia aparece como conceito chave que conecta, via teoria dos ensembles, as configurações microscópicas de um sistema com seu comportamento macroscópico. Por sua vez, o limite termodinâmico que assegura a existência desta conexão micro \leftrightarrow macro, assenta-se sobre o conceito fundamental de extensividade da energia e da entropia.

Contrário ao senso comum, o limite termodinâmico [3] não é imprescindível à definição da mecânica estatística microcanônica ou à descrição de transições de fase [4, 5]. Este preconceito consolidou-se com o uso de ensembles derivados do microcanônico, como o canônico e o grande canônico, cujas formulações necessitam tacitamente do limite termodinâmico [5]. Como a abordagem microcanônica baseia-se na exata contagem de microestados, ela é ideal ao estudo de sistemas "pequenos" [2, 6, 7]. O termo "pequenos" designa aqui sistemas que interagem via forças de longo alcance e, ou tem poucos graus de liberdade. Dentre os quais encontram-se os sistemas gravitacionais [8], de spin [9, 10, 11, 12, 13, 15, 16], plasmas [17, 18, 19, 20, 21, 22] e biomoleculares como as proteínas.

Proteínas [23, 24] são heteropolímeros de elevada massa molecular, compostas por um grande número de aminoácidos de até 20 diferentes tipos [24, 25]. Dentre as principais funções protéicas, que apresentam alta especificidade em sistemas biológicos vivos, estão a estrutural e a metabólica. Essas características devem-se em grande parte à estrutura geométrica tridimensional, também chamada de terciária, que caracteriza cada proteína e confere-lhe utilidade biológica [24]. Ao processo que conforma espacialmente uma cadeia de

aminoácidos em sua estrutura nativa denomina-se enovelamento (no inglês, *foldings*). Suas origens físicas estão nas complexas interações atômicas, que produzem perfis de energias livres rugosos e afunilados [26]. Atualmente, potenciais interatômicos [27, 28, 29, 30] de origem eletromagnética [31] já conseguem reproduzir certas estruturas nativas via simulações *ab initio*.

Sabemos ainda que em determinados momentos da síntese protéica [23] podem ocorrer defeitos conformacionais (*misfoldings*) que degradam funcionalmente as proteínas afetadas. Esses casos isolados podem repetir-se originando agregados protéicos [32, 33, 34] ricos em hélices beta-cruzadas. Este tipo de estrutura é conhecida por sua tenacidade e habilidade em induzir doenças degenerativas. Dentre essas doenças estão as neurodegenerativas, um termo que designa perda progressiva de estruturas e funções neuronais, levando à morte.

A neurodegeneração pode acontecer em diferentes níveis neurais, variando dos níveis molecular ao sistêmico. Vários males, como o de Parkinson [35], Alzheimer [36] e Huntington [37], são proteinopatias relacionadas a agregação protéica em nível subcelular. Normalmente essas doenças não são transmissíveis por um vetor etiológico. Contudo, os males da "vaca louca" e de Creutzfeldt-Jakob (DCJ) [38, 39] são causados por proteínas "infecciosas" e auto-replicas chamadas Prions [40, 41], que violam o dogma central da biologia [42]. Do ponto de vista físico, existem muitas similaridades no desenvolvimento de diversas proteinopatias [32, 34]. Uma melhor compreensão destes aspectos pode ser alcançada por meio de simulações computacionais, aliadas à modelagem físico-estatística [26, 43, 44, 45], as quais poderão auxiliar no desenvolvimento de novas terapias.

A presente tese procura investigar o comportamento de algumas proteínas de grande interesse biológico, como é o caso dos domínios Src SH3 [48] e das Prions humanas [40, 41]. Os métodos que utilizamos, como os ensembles generalizados [49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62] e as simulações de Monte Carlo [63, 64], permitem calcular diretamente a entropia microcanônica. Deste modo, é possível constatar profundas semelhanças físicas no comportamento de proteínas e certos modelos de spin, classificados como "pequenos". Este é o caso do modelo de spin de Blume-Capel de alcance infinito, que estudamos nesta tese como laboratório teórico, por causa do seu peculiar comportamento termodinâmico. A abordagem microcanônica mostra-se valiosa nestas situações, pois permite formalizar e descrever unificadamente transições de fase mesmo em sistemas finitos, como no caso biológico do enovelamento e da agregação protéicos [43, 44, 45, 46, 47].

Visando minimizar as exigências de poder computacional adotamos potenciais interatômicos simplificados, do tipo grão grosso (*coarse-grained*) [46, 65, 66, 67, 68, 69, 70, 71], ao invés dos poderosos e complexos CHARM e AMBER [27]. Pela sua simplicidade, escolhemos o chamado modelo AB [68, 69], cuja literatura reporta apenas aplicações no estudo de seqüências artificiais, as Fibonacci [43, 44, 46].

No modelo AB as principais interações consideradas são de caráter efetivo, e descrevem o comportamento hidrofóbico-hidrofílico dos aminoácidos [72] envolvidos. As interações são modeladas por meio de pseudo-átomos, o que reduz notavelmente o número de graus de liberdade do sistema, conferindo ao modelo grande leveza computacional. Por sua vez, a interação interprotéica, responsável pela biologia da agregação e dos diversos tipos de *misfoldings*, é modelada por um termo hamiltoniano [46] que considera pares de aminoáci-

dos. Nossos resultados mostraram-se promissores, revelando similaridades com observações experimentais [34, 73, 74] e estudos teóricos prévios [48].

O texto é organizado como segue. No Capítulo 2 introduzimos a termoestatística microcanônica segundo a formulação Boltzmanniana [1, 3] e apresentamos o formalismo elaborado por Gross [2] e generalizado por Kastner [75, 76] para descrever transições de fase microcanônicas. Note que o limite termodinâmico não é exigido por este formalismo. Por fim, abordamos alguns ensembles generalizados como o multicanônico [61] e o gaussiano estendido (EGE) [51, 56], os quais fornecem robustas regularizações microcanônicas e formas alternativas para estimar a entropia [62, 77].

No Capítulo 3 abordamos as simulações de Monte Carlo que em mecânica estatística são importantes ferramentas numéricas. Revisamos os métodos de Monte Carlo estáticos e dinâmicos, como o algoritmo de Metropolis [64], além de descrevermos como estimar erros estocásticos via rigorosas estimativas da autocorrelação [78]. Introduzimos em seguida o algoritmo multicanônico (MUCA) [61, 62] que determina observáveis canônicos via repesagem [63] e microcanônicos via os próprios parâmetros multicanônicos.

Sendo esta uma tese interdisciplinar apresentamos no Capítulo 4 uma revisão sobre proteínas no que se refere à sua constituição química [23], funcionalidade e taxonomia [24]. Revisitamos o conceito de proteinopatias enquanto doenças degenerativas resultantes do mal enovelamento e agregação protéicos subcelulares. Além disso, a modelagem teórica desses sistemas biomoleculares é enfocada pela apresentação do modelo AB, considerado o mais propício aos nossos propósitos.

No Capítulo 5 utilizamos o ensemble gaussiano estendido para proceder a uma nova resolução analítica do modelo Blume-Capel de alcance infinito [13, 79, 80, 81, 82]. Este modelo de spin 1 é semelhante ao de Ising, possuindo contudo interações competitivas de alcance infinito. A literatura moderna reporta a inequivalência entre suas conhecidas soluções nos ensembles canônico e microcanônico [9]. Todavia, uma solução em um ensemble interpolante era até agora desconhecida. Nossa nova solução [83] além de recuperar os resultados anteriores como casos limite, também evidencia as metaestabilidades canônicas. Efetuamos por fim um estudo dos pontos tricríticos do modelo BC para ilustrar propriedades da abordagem EGE.

O capítulo 6 dedica-se às simulações microcanônicas de proteínas propriamente ditas. Os exemplares que investigamos são cadeias de Fibonacci, o domínio Src SH3 (código PDB: 1NLO) e a Príon humana (código PDB: 1HJM). As cadeias Fibonacci, que foram artificialmente desenhadas [46], são simuladas para efeito de comparação com proteínas reais. Os peptídeos 1NLO e 1HJM foram mapeados em sequências do tipo AB, com as quais efetuamos extensas simulações para obter a sua termodinâmica microcanônica pela análise dos parâmetros multicanônicos. Nossos resultados numéricos mostram-se coerentes e indicam a existência de um comportamento biológico universal. Além disso, constatamos semelhanças mecânico-estatísticas típicas de sistemas "pequenos" entre as transições de enovelamento, de agregação e as exibidas pelo modelo BC resolvido no EGE.

O Capítulo 7 conclui esta tese com discussões finais e o delineamento de perspectivas futuras. Apontamos aqui novas questões surgidas neste estudo e vias para investigá-las. Certos tópicos físicos e matemáticos são discutidos à parte como apêndices em anexo.

Capítulo 2

Termostatística microcanônica

... By the study of Boltzmann I have been unable to understand him. He could not understand me on account of my shortness, and his length was and is an equal stumbling-block to me. Hence I am very inclined to join the glorious company of supplanters and to put the whole business in about six lines ...

[J.C. MAXWELL a P.G. Tait, agosto de 1873]

A mecânica estatística permitiu uma compreensão da fenomenologia descrita pela termodinâmica a partir de abordagens microscópicas. Inicialmente concebida para explicar fenômenos simples, como o comportamento de gases monoatômicos, interagindo via potenciais de curto alcance, essa área da física desenvolveu-se rápido e atualmente é crucial para o entendimento de sistemas complexos. Entretanto, para assegurar sua ampla empregabilidade é preciso garantir a existência do limite termodinâmico, de modo que todas as formulações mecânico-estatísticas, dadas pela teoria de ensembles, sejam equivalentes [3].

Por outro lado, nos casos em que o limite termodinâmico não se aplica, como ocorre com sistemas ditos "pequenos" [2, 6], cujo alcance das interações equivale ao tamanho do sistema, surge o fenômeno da inequivalência de ensembles. Nesta situação as previsões físicas efetuadas dependem do formalismo estatístico adotado. Dado tal grau de arbitrariedade, que é antes formal do que realmente físico, resta entender qual das abordagens mecânico-estatísticas é a relevante.

Neste capítulo revisamos a abordagem microcanônica usual *à la Boltzmann* [2, 3, 6], pois dela se deduzem, via teoria das transformações de Laplace e Legendre, todas as demais¹. Este formalismo fornece ainda uma prescrição simples e direta para a caracterização das transições de fase, sendo aplicável mesmo a sistemas "pequenos". Por fim, introduzimos noções sobre ensembles generalizados, como o gaussiano estendido e o multicanônico. Estes ensembles são universalmente equivalentes ao microcanônico e importantes para simulações numéricas.

¹Nos casos em que há inequivalência de ensembles as transformações de Laplace não são inversíveis, pois ocorrem regiões não côncavas na entropia. Ainda assim os ensembles canônico e grande canônico são dedutíveis a partir do microcanônico, mas não o inverso!

2.1 O ensemble microcanônico

A formulação microcanônica surgiu das investigações do físico austríaco Ludwig Boltzmann [1] visando descrever o comportamento cinético dos gases a partir de um ponto de vista microscópico. O conceito termodinâmico de entropia é fundamental nessa abordagem e a ele foi associado, em nível microscópico, a idéia de desorganização estatística de sistemas físicos isolados. O equilíbrio termodinâmico nesse contexto é realizado pela maximização da entropia do sistema.

Define-se entropia microcanônica $S(E, N, V)$, que para um sistema mecânico extensivo dependerá da energia E , do número de partículas N e do volume V , como

$$S(E, N, V) = k_B \ln W(E, N, V), \quad (2.1)$$

em que k_B é a constante de Boltzmann e W é a função de partição microcanônica². Ou seja, W representa o número de microestados acessíveis e compatíveis com um dado macroestado termodinâmico, cuja ocupação é equiprovável microcanonicamente em sistemas ergódicos. Por causa da propriedade de equiprobabilidade dos estados a abordagem microcanônica não exhibe barreiras de probabilidade nas vizinhanças de transições de fase.

O formalismo microcanônico incorpora naturalmente uma prescrição para o cálculo de $W(E, N, V)$ através de um processo de particionamento do espaço de fase. Como exemplo, consideremos um sistema de N corpos cuja energia total é fixada em E . Se sua dinâmica for regida pela hamiltoniana H_N , pode-se obter W calculando-se a integral vinculada,

$$W = \epsilon_0 \int \frac{1}{N!} \left[\frac{d^3 p d^3 q}{h^3} \right]^N \delta(E - H_N(p, q)), \quad (2.2)$$

em que as constantes ϵ_0 e h (de Planck) tem dimensões apropriadas.

De fato, como a abordagem microcanônica é mecanicamente bem definida, mesmo em um nível microscópico, os vínculos do sistema são impostos a cada um dos membros do ensemble, ou seja, a cada ponto no espaço de fase. Por isso, a formulação microcanônica tem sentido físico mesmo para sistemas "pequenos", e independentemente da existência do limite termodinâmico. Este aspecto contrasta, por exemplo, com a abordagem grande canônica. Esta abordagem se deduz da microcanônica via transformada dupla de Laplace [2, 3] em que os mecanismos de troca de energia e partículas necessitam de acoplamentos com banhos térmicos infinitos (i.e. reservatórios).

Fica claro que uma das principais virtudes da abordagem microcanônica é sua habilidade em descrever diretamente e de modo estatístico, o comportamento global de sistemas de muitos-corpos empregando apenas alguns poucos parâmetros mecânicos de controle (e.g. E, N, V). Ainda, a entropia definida por Boltzmann é, em nível clássico, uma função contínua, multiplamente diferenciável com relação à energia e côncava³ globalmente⁴ no limite termodinâmico.

²Do alemão *Wahrscheinlichkeit*: probabilidade.

³Formalmente, uma função real f definida em um intervalo é dita côncava, se para quaisquer dois pontos x_1 e x_2 em seu domínio C , e para qualquer t em $[0, 1]$, cumpre-se $f(tx_1 + (1-t)x_2) \geq tf(x_1) + (1-t)f(x_2)$. Em adição, $f(x)$ é côncava em $[a, b]$ se e somente se a função $-f(x)$ é convexa em $[a, b]$.

⁴Embora admita localmente os ditos *intrusos convexos* [2]

Por sua vez, a conexão do formalismo microcanônico com a termodinâmica se dá pela entropia, cujas razões entre taxas de variação fornece observáveis termodinâmicos como a temperatura ou o calor específico. Como ilustração, consideremos um sistema magnético descrito pela hamiltoniana $H_N(m)$, em que o número de partículas N , a energia por partícula $e = H_N/N$ e a magnetização por partícula $m = M/N$ [2] são utilizados para obter:

- Temperatura (T)

$$\frac{1}{T(e)} \doteq \beta(e) = \frac{\partial}{\partial e} s(e, m), \quad (2.3)$$

- Calor específico (c_V)

$$c_V(e) \doteq \frac{de}{dT} = -\frac{s_{mm}}{T^2 d(e, m)}, \quad (2.4)$$

- Susceptibilidade magnética específica (χ_T)

$$\chi_T \doteq \left(\frac{\partial m}{\partial B} \right)_T = -\frac{s_{ee}}{d(e, m)}. \quad (2.5)$$

Fizemos uso na Eq. (2.5) da curvatura gaussiana. Esta curvatura corresponde ao determinante da matriz hessiana da entropia $d[s(e, m)]$, definida como

$$d_S(e, m) = \det \begin{pmatrix} s_{ee} & s_{em} \\ s_{me} & s_{mm} \end{pmatrix}. \quad (2.6)$$

2.2 Transições de fase microcanônicas

Transições de fase são usualmente estudadas em mecânica estatística pela abordagem de Lee e Yang [4], em que os chamados zeros complexos da função de partição grande canônica são analisados. Uma revisão detalhada dessa abordagem foge ao escopo desta tese, mas é importante notar que nesse esquema inexitem transições de fase em sistemas finitos⁵. Entretanto, é crucial aos nossos propósitos entender se na ausência do limite termodinâmico inexitem de fato transições de fase, ou se este efeito é apenas um artefato de um particular formalismo.

Consideramos que quando há equivalência de ensembles a descrição física de transições de fase é naturalmente independente do formalismo adotado. Mas nos casos em que há inequivalência de ensembles a formulação estatística é determinante quanto à fenomenologia observável. No caso microcanônico, cuja conexão termodinâmica se dá via entropia de Boltzmann, é o comportamento desta grandeza que discrimina a natureza das transições de fase [2, 6, 7].

⁵Isso porque fora do limite termodinâmico tem-se N finito, e portanto a função de partição Z pode ser escrita como uma soma finita e analítica de $(z = e^{\mu/T})^N$ termos. Para completar o raciocínio, devemos lembrar que o grande potencial é $\propto \frac{1}{V} \ln Z$ e também é analítico se Z o for.

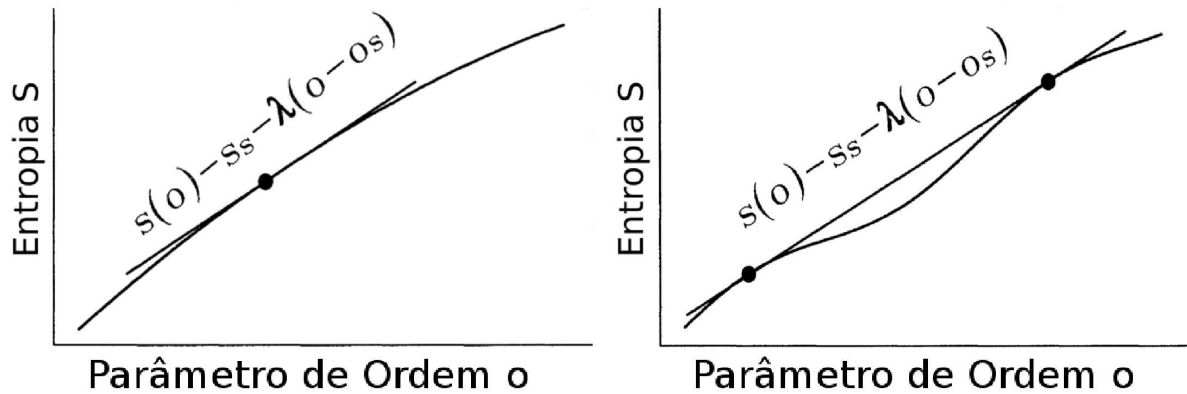


Figura 2.1: Esquerda: situação que exemplifica uma entropia côncava, i.e., fase única. Direita: ilustração de um intruso convexo na entropia, como o que ocorre nas transições canônicas de primeira ordem; aqui os dois pontos-de-sela correspondem a duas fases distintas. Adaptado da referência [2].

De forma geral, transições de fase microcanônicas são definidas pelos: “*pontos e regiões de curvatura não negativa da hipersuperfície entrópica $S_N(P_1, \dots, P_i)$, descrita no espaço de fase em função das quantidades mecânicas conservadas e extensivas $\{P_1, \dots, P_i\}$ das quais S_N depende, como a energia, massa, magnetização, momentum angular, etc*” [2, 6, 7]. Portanto, para uma rigorosa utilização destes conceitos define-se a curvatura, ou matriz hessiana (H_S), da entropia $S_N(P_1, \dots, P_i)$ como

$$H_{S(P_1, P_2, \dots, P_i)} = \begin{pmatrix} \partial_{P_1} \partial_{P_1} S & \dots & \partial_{P_1} \partial_{P_i} S \\ \vdots & \ddots & \vdots \\ \partial_{P_i} \partial_{P_1} S & \dots & \partial_{P_i} \partial_{P_i} S \end{pmatrix}. \quad (2.7)$$

Para caracterizar devidamente as regiões de transição de fase utilizam-se técnicas capazes de extrair invariantes geométricos e algébricos da hipersuperfície S_N . Por exemplo, a geometria diferencial [5] fornece-nos a *curvatura gaussiana* (d_S), i.e. o determinante de H_S , que é um importante invariante [2]. Podemos ainda expressá-la em termos dos autovalores ordenados $\{\lambda_1, \dots, \lambda_N\}$ da curvatura entrópica como

$$d_S = \det [H_{S(P_1, \dots, P_i)}] = \lambda_1 \lambda_2 \dots \lambda_N. \quad (2.8)$$

Assim, todos os possíveis comportamentos fenomenológicos conhecidos nas abordagens canônica ou grande canônica enquadram-se nos seguintes casos:

- **Uma única fase estável:** é observada quando $d_S > 0$ e $\lambda_1 < 0$. Nessa situação S_N é côncava em todas as direções no limite termodinâmico. Temos aqui um mapeamento bi-unívoco entre as grandezas termodinâmicas computadas via quaisquer ensembles. Vide Figura (2.1).

- **Transição de fase de primeira ordem:** nesse caso observa-se separação de fases e tensão interfacial e é caracterizada por $d_S < 0$ e $\lambda_1 > 0$. Aqui S_N possui um intruso convexo (curvatura "para cima", Figura 2.1) na direção do autovetor v_{λ_1} associado à componente de maior curvatura λ_1 . Existem aqui dois pontos-de-sela: naquele mais à esquerda o sistema é totalmente "líquido" e no outro é "gasoso". Toda a região convexa da entropia é mapeada em um único ponto no ensemble grande canônico; portanto, se a curvatura de S_N for, por exemplo, igual a $\lambda_1 \geq 0$ haverá *inequivalência de ensembles*⁶. Surgem valores negativos das funções resposta, como no calor específico, como se pode constatar através das definições nas Eqs. (2.4) e (2.5), Figura (2.1).
- **Transição de fase de segunda ordem ou contínua:** trata-se do tipo de transição de fase em que desaparece a tensão interfacial, ambas as fases vizinhas tornam-se indistinguíveis. Neste caso as linhas críticas são aquelas onde $d_S = 0$ e $\vec{v}_{\lambda=0} \cdot \vec{\nabla} d_S = 0$, em que $\vec{v}_{\lambda=0}$ é o autovetor de H_S associado ao autovalor $\lambda = 0$ de maior curvatura. Nessas situações podem ocorrer as conhecidas catástrofes da transformada inversa de Laplace⁷ $E \rightarrow T$.
- **Pontos multicríticos:** ocorrem em regiões em que mais de duas fases tornam-se indistinguíveis; estão associados a locais em que ocorrem divisões das linhas de um diagrama de fases. Matematicamente são descritos por $d_S = 0$ e $\vec{\nabla} d_S = 0$.

2.3 Inequivalência de ensembles e transições de fase

A grandeza termodinâmica que caracteriza transições de fase no formalismo canônico da mecânica estatística é a energia livre de Helmholtz (F). Ela é obtida da energia interna U e da entropia S pela transformação de Legendre⁸ $F = U - TS$, vide apêndice C. Para

⁶Sob outro aspecto, vale notar que a função de partição grande canônica é definida via dupla transformada de Laplace da densidade de estados microcanônica [i.e., de $\Omega(e, n, V) = e^{s(e, n, V)}$] como

$$\Xi(\mu, T, V) \doteq e^{-\beta F(\mu, T, V)} = \frac{V^2}{\epsilon_0} \int_0^\infty de \int_0^\infty dne^{-V[e - \mu n - Ts(e, n, V)]/T},$$

de onde pode-se mostrar [2] que assintoticamente

$$\frac{F(\mu, T, V)}{V} \rightarrow e - \mu n - Ts + \frac{T \ln \sqrt{d_S}}{V} + O\left(\frac{\ln V}{V}\right).$$

Portanto, para $d_S > 0$ a energia livre específica tende ao limite termodinâmico a seu valor típico $f \rightarrow e - \mu n - Ts$. Entretanto, se $d_S = 0$ ocorrem divergências em F , mesmo para sistemas finitos, e para $d_S < 0$ inexistente uma definição da energia livre!

⁷Por exemplo, nessa situação a transformada de Laplace que converte a densidade de estados $\Omega = \Omega(e)$ na função de partição canônica: $Z(\beta) = \int_0^\infty e^{-\beta e} \Omega(e) de$ não é inversível para todo β pois a curva $e \times \beta(e)$ tem loops em formato de S. Logo, falha a descrição fenomenológica de transições de fase em termos de variáveis intensivas, como $T(e)$ que necessita de construções auxiliares como a de Maxwell sobre os loops de Van der Waals.

⁸Este tipo de transformada, e sua inversa, sempre retornam funções côncavas.

que esta transformação e sua inversa façam sentido, S precisa ser localmente côncava, o que termodinamicamente esta associado ao critério de estabilidade deste ensemble⁹. Uma vez conhecida F a análise das suas divergências e de suas funções resposta, como função de um parâmetro intensivo (por exemplo, a temperatura), permite classificar fenomenologicamente [5] as transições de fase. Assim, nesta abordagem as transições de fase são ditas de $n - \text{ésima}$ ordem se a derivada $n - \text{ésima}$ de F for descontínua¹⁰, ao passo que suas funções resposta são positivas definidas¹¹.

Por outro lado, segundo o formalismo microcanônico, transições de fase são classificadas [2, 75] pelo comportamento da entropia S e de suas funções resposta com relação a um parâmetro extensivo como a energia U . Por exemplo, a Eq. (2.3) implica que transições de fase de primeira ordem neste formalismo exibem descontinuidades na temperatura. Como é a maximização de S que garante a estabilidade termodinâmica nesta abordagem, e não a concavidade local de S como no caso canônico, ambos ensembles podem exibir (em princípio) fenomenologias bastante distintas. Este é o caso quando S for localmente convexa, o que implica em funções resposta microcanônicas negativas segundo as Eq. (2.4) e Eq. (2.5) [14].

Normalmente a escolha de determinado ensemble é uma questão de conveniência técnica, pois no limite termodinâmico os resultados físicos independem do formalismo adotado. Existem contudo situações em que resultados, como a ordem da transição de fase, diferem e originam inequivalência de ensembles. Isto pode ocorrer apenas quando há transições de fase canônicas de primeira ordem em sistemas que interagem via potenciais de longo alcance¹², como nos gravitacionais [8, 13]. Nestes sistemas a entropia microcanônica apesar de ser contínua exibe regiões convexas. Ao leitor interessado apontamos [2, 13, 14], que

⁹Situações em que S não é côncava implicam metaestabilidades termodinâmicas no ensemble canônico, como ocorre em transições de fase de primeira ordem. Nestes casos é usual aparecerem *loops* de Van der Waals nos diagramas de fase do sistema, o que exige construções à la Maxwell para reproduzir a curva dos calores latentes [3].

¹⁰Nas vizinhanças de uma transição de fase contínua, ou de segunda ordem, há uma temperatura crítica T_c em que as funções resposta de F como c e χ apresentam divergências da forma

$$\begin{aligned} c &\propto |T - T_c|^{-\alpha} \\ \chi &\propto |T - T_c|^{-\gamma}. \end{aligned} \quad (2.9)$$

O grau de divergência ditado pelos chamados expoentes críticos, como α e γ , permite classificar em classes de universalidade o comportamento físico destes sistemas durante transições de fase. Por outro lado, em transição de fase de primeira ordem tais expoentes críticos não se associam ao conceito de universalidade [5, 3].

¹¹Considere por exemplo o caso do calor específico calculado com o formalismo canônico,

$$c_V \doteq \frac{dU}{dT} = -\beta^2 \frac{\partial \langle E \rangle}{\partial \beta} = \beta^2 \frac{\partial \ln(Z)}{\partial \beta} = \beta^2 \left(\langle E^2 \rangle - \langle E \rangle^2 \right) \geq 0,$$

onde $\beta = 1/T$ é o inverso da temperatura física T de equilíbrio do sistema acoplado ao reservatório térmico. Aqui a energia média canônica $\langle E \rangle_\beta$ é termodinamicamente identificada com a energia interna do sistema U .

¹²É usual definirem-se potenciais de longo alcance [13] como aqueles que dependam da distância r como $r^{-\alpha}$, sendo $\alpha < d$.

justificam e exemplificam a fundamentalidade microcanônica.

Para efeito de ilustração consideremos duas situações de transições de fase em que as descrições oferecidas por ambos os ensembles podem apresentar inequivalências:

- Sempre que as transições de fase microcanônicas forem de primeira ordem, ou seja, descontínuas nas curvas calóricas (i.e., $\beta \times E$ tendo saltos), elas serão de primeira ordem também no ensemble canônico. Apesar de neste caso haver concordância na ordem da transição de fase segundo ambos os ensembles, a termodinâmica obtida a partir das duas abordagens será inequivalente. Isto porque transições de fase microcanônicas de primeira ordem apresentam normalmente funções resposta negativas.
- Quando as transições de fase forem de primeira ordem na abordagem canônica, o que se reflete na presença de uma curva calórica (microcanônica) $\beta \times E$ contínua mas dotada de *loops* de Van der Waals, haverá possíveis inequivalências. Nesta situação há *loops* na referida curva calórica, que são canonicamente alterados (i.e., aplainados) a lâ Maxwell, e são nestas regiões que surgem calores específicos microcanônicos negativos e divergentes, o que sinaliza transições de segunda ordem neste ensemble.

Devemos ainda mencionar que nos casos em que a transição de fase é crítica, ou de segunda ordem canônica, há equivalência total de ensembles.

2.4 Ensembles generalizados

Apresentaremos nas subseções seguintes dois ensembles generalizados, que são extremamente úteis para a estimativa tanto analítica quanto numérica da entropia microcanônica como definida na Eq. (2.2) [2, 6, 7].

No primeiro caso temos o ensemble multicanônico, que surgiu sob inspiração puramente algorítmica para driblar a degradação de desempenho que simulações numéricas enfrentam ao redor de transições de fase [61, 62]. Nele, uma estimativa do tipo *piecewise* para a entropia microcanônica é implementada através de parâmetros ditos multicanônicos.

O segundo ensemble por sua vez é conhecido como gaussiano estendido (EGE) [50, 51, 52, 54, 55, 56, 58] e comporta-se como um ensemble interpolante entre o microcanônico e o canônico [56]. Esta situação interpolante descreve sistemas acoplados a banhos térmicos finitos. Recentemente, foi mostrado ainda que o EGE é equivalente aos ensembles microcanônico e MUCA [17, 19, 20, 21, 22, 49, 57, 59, 60].

2.4.1 O ensemble multicanônico

Um considerável avanço na determinação microcanônica das densidades de estado data a 1991 com a introdução do chamado ensemble multicanônico [61, 62]. Recordemos que na abordagem canônica tradicional o sistema permanece em contato com um reservatório com temperatura fixa $T = 1/k_B\beta$, e tem as energias E_k da configuração k descritas pelo peso de Boltzmann-Gibbs

$$w_B(E_k) = e^{-\beta E_k}. \quad (2.10)$$

Enquanto os estados com energia E são distribuídos com probabilidade

$$P_B(E) = c_\beta w_B(E) = c_\beta \Omega(E) e^{-\beta E}, \quad (2.11)$$

onde a constante de normalização c_β é introduzida para garantir que $\sum_E P_B(E) = 1$.

Como a densidade de estados $\Omega(E)$ é uma função que cresce rapidamente, enquanto o fator de Boltzmann decai exponencialmente com E , temos que $P_B(E)$ tem geralmente a forma de uma gaussiana ou apresenta picos duplos [3]. No caso de uma transição de fase de primeira ordem, o ponto crítico $\beta_c(L)$ em um sistema de volume finito L^d é definido de forma que a distribuição de energia $P_B(E, L)$ apresente dois picos de alturas iguais nas energias E_{max}^1 e E_{max}^2 , $P_B(E_{max}^1, L) = P_B(E_{max}^2, L)$. Entre estes dois valores ocorre a energia E_{min} , correspondendo ao mínimo de $P_B(E, L)$ [63].

Sabemos que as configurações em E_{min} são exponencialmente suprimidas segundo

$$P_{min} = P(E_{min}) = c_f L^p \exp(-f^s A), \quad (2.12)$$

onde f^s é a tensão interfacial e $A = 2L^{d-1}$ é a área entre as duas fases para uma rede L^d . Temos ainda as constantes c_f e p ($p = d - 1$).

Entretanto, ainda que lidando com a mecânica estatística canônica, o peso de Boltzmann não é necessariamente uma prescrição computacionalmente adequada para todos os casos. Numericamente, este peso não seleciona configurações representativas da interface em transições de primeira ordem [63].

Uma solução é o uso do ensemble multicanônico [61, 62], que foi inicialmente projetado para calcular a tensão interfacial em simulações no ensemble canônico de Gibbs. Ele foi desenvolvido empiricamente com a finalidade de produzir configurações representativas das fases do sistema e com a exigência de eficientemente ultrapassar as barreiras de energia livre. Procurou-se então amostrar, em um intervalo apropriado da energia, configurações geradas com o seguinte peso

$$w_{muca} = e^{-b(E_k)E_k + a(E_k)}, \quad (2.13)$$

ao invés do tradicional peso de Boltzmann-Gibbs. O objetivo era obter uma nova distribuição de probabilidades, com densidade de estados $n(E)$ multicanônica

$$P_{muca}(E) = c_{muca} n(E) w_{muca}(E) \approx \text{constante}, \quad (2.14)$$

que não fosse fortemente concentrada, como no caso usual da Eq. (2.11).

Essa nova distribuição faz com que na região ao redor de P_{min} não ocorra mais a supressão de configurações, logo, o sistema passaria a visitar igualmente todas as configurações disponíveis nesse intervalo de energias. A nova função $b(E)$ é interpretada nesse esquema como uma temperatura microcanônica na energia E e $a(E)$ passa a ser uma espécie de fugacidade. A distribuição canônica original $P(E)$ pode ser obtida [63] por meio de uma repesagem, isto é

$$P(E) = \frac{P_{muca}(E)}{c_{muca} w_{muca}(E)} c_\beta e^{-\beta E}. \quad (2.15)$$

Esta relação é rigorosa pois os pesos $w_{muca}(E)$ já são conhecidos nesta etapa.

Percebeu-se posteriormente que os pesos procurados $w_{muca}(E_k)$ correspondem a uma boa aproximação para os pesos microcanônicos $w_{1/\Omega(E)}(E_k)$, isto é

$$w_{muca}(E_k) \approx w_{1/\Omega(E)}(E_k) = \frac{1}{\Omega(E_k)} \quad (2.16)$$

como consequência direta da Eq. (2.14). Ou seja, aqui a entropia microcanônica é diretamente estimada como uma função *piecewise* do tipo $S(E_k) = b_k(E_k)E_k - a_k(E_k)$.

Contudo, há uma dificuldade inicial em aplicar o algoritmo multicanônico para estimar a densidade de estados microcanônica visto que os pesos na Eq. (2.16) são *a priori* desconhecidos. Portanto, para estimar adequadamente o conjunto de pesos $\{a_k, b_k\}$ é preciso utilizar relações de recorrência em simulações preliminares sucessivas¹³, que antecedem a simulação produtiva propriamente dita. Uma vez fixados os $w_{muca}(E_k)$, a simulação é efetuada segundo métodos usuais [64].

Por fim, o cálculo de grandezas termodinâmicas descritas pelo ensemble canônico pode ser obtido do ensemble multicanônico via repesagem dos dados provindos de sua série temporal. Por exemplo, a energia média canônica a uma temperatura $1/\beta$ é calculada a partir das medidas multicanônicas E_i ,

$$\bar{E}(\beta) = \frac{\sum_{i=1}^n E_i w_{muca}^{-1}(E_i) e^{-\beta E_i}}{\sum_{i=1}^n w_{muca}^{-1}(E_i) e^{-\beta E_i}}. \quad (2.17)$$

2.4.2 O ensemble gaussiano estendido

O ensemble gaussiano surgiu nos anos 80 com a finalidade de acelerar os métodos de Monte Carlo usuais [50, 51, 52, 54, 58]. Posteriormente, este ensemble foi reinterpretado como um esquema regularizador para o ensemble microcanônico [55]. Ele interpola por meio de um parâmetro γ , relacionado à capacidade calorífica de um banho térmico finito, a física dos ensembles canônico e microcanônico [13, 83]. Uma generalização ulterior deste esquema produziu o ensemble gaussiano estendido (EGE) [47, 56, 77]. Recentemente foi demonstrado [57] haver equivalências entre os ensembles microcanônico, gaussiano estendido, multicanônico e de Tsallis em certos regimes termodinâmicos.

Deduzimos aqui as propriedades do EGE [56, 83] utilizando métodos comumente encontrados na literatura para obter o ensemble canônico [3] a partir do microcanônico, por exemplo, como é ilustrado no Apêndice E.

Inicialmente, consideremos um sistema a com energia E e entropia S , acoplado a um banho térmico b com energia E_b e entropia S_b que troca energia com a . Logo, a energia total do sistema isolado será $E_t = E + E_b$ e sua entropia total é S_t . Neste caso o equilíbrio

¹³Geralmente utilizam-se relações de recorrência entre os parâmetros b_n e a_n

$$a_n(E - \epsilon) = a_n(E) + [b_n(E - \epsilon) - b_n(\epsilon)]E,$$

$$b_{n+1}(E) = b_n(E) + [\ln H_n^{muca}(E + \epsilon) - \ln H_n^{muca}(E)]/\epsilon$$

provenientes das n -ésimas simulações sucessivas [61, 62]. Vide próximo capítulo.

térmico é alcançado quando S_t é máxima e a energia E do sistema a flutua ao redor de um valor médio U de equilíbrio. Então, a energia mais provável é tal que uma expansão da entropia do banho térmico S_b , ao redor do equilíbrio $E_t - U$, resulta em segunda ordem

$$S_b(E_b) = S_b(E_t - U) + \left(\frac{dS_b}{dE_b} \right)_{E_t - U} (U - E) + \frac{1}{2} \left(\frac{d^2S_b}{dE_b^2} \right)_{E_t - U} (U - E)^2 + \dots \quad (2.18)$$

Se considerarmos que estas derivadas dependem das propriedades físicas do banho térmico, é conveniente definirmos

$$\left(\frac{dS_b}{dE_b} \right)_{E_t - U} = \alpha, \quad (2.19)$$

e

$$\frac{1}{2} \left(\frac{d^2S_b}{dE_b^2} \right)_{E_t - U} = -\gamma. \quad (2.20)$$

No caso de haver um banho térmico infinito, representado por um reservatório, estaríamos trabalhando no ensemble canônico tradicional, ou seja $\alpha = \beta = 1/(k_B T)$ e $\gamma = 0$. No limite oposto, em que há um banho térmico infinitesimal, temos o caso microcanônico com $\gamma \rightarrow \infty$ e $E_t \equiv U$.

O método dos multiplicadores de Lagrange [3, 56] nos dá o peso gaussiano estendido

$$w_{EGE} = e^{-\alpha E - \gamma(E-U)^2}, \quad (2.21)$$

e a densidade de probabilidade para o EGE

$$P_{\gamma, \alpha}(E) = \frac{\rho(E) e^{-\alpha E - \gamma(E-U)^2}}{Z_\gamma(U, \alpha)}, \quad (2.22)$$

com a qual define-se a função de partição deste ensemble,

$$Z_\gamma(U, \alpha) = \int \rho(E) e^{-\alpha E - \gamma(E-U)^2} dE. \quad (2.23)$$

Daqui, define-se o potencial termodinâmico generalizado,

$$\Phi_\gamma(U, \alpha) = -\ln Z_\gamma(U, \alpha). \quad (2.24)$$

Enquanto o parâmetro U pode ser determinado autoconsistentemente pela seguinte relação

$$U \doteq \left(\frac{\partial \Phi_\gamma}{\partial \alpha} \right) = \int E P_{\gamma, \alpha}(E) dE. \quad (2.25)$$

Se aplicarmos a transformação de Legendre-Fenchel (LF) [3] ao potencial $\Phi_\gamma(U, \alpha)$ encontraremos a entropia generalizada S_γ do EGE,

$$S_\gamma(U) = \alpha \left(\frac{\partial \Phi_\gamma}{\partial \alpha} \right)_\gamma + \gamma \left(\frac{\partial \Phi_\gamma}{\partial \gamma} \right)_\alpha - \Phi_\gamma. \quad (2.26)$$

Vale notar também que, como ocorre no caso da termodinâmica canônica, pode-se definir um calor específico generalizado para o sistema [51, 52]. Este calor específico é dependente de γ ,

$$C_\gamma \doteq -\alpha^2 \left(\frac{\partial U}{\partial \alpha} \right)_\gamma = \frac{\langle (E - U)^2 \rangle}{1 - 2\gamma \langle (E - U)^2 \rangle}. \quad (2.27)$$

Ressaltamos aqui que a positividade das flutuações $\langle (E - U)^2 \rangle$ não implica necessariamente na positividade de C_γ .

Capítulo 3

Simulações

God does not care about our mathematical difficulties. He integrates empirically
[A. EINSTEIN a L. Infield, 1942].

A obtenção da entropia microcanônica partindo-se diretamente da definição na Eq. (2.2) é uma tarefa nem sempre possível analiticamente. O que reforça a necessidade de métodos numéricos como as simulações de Monte Carlo (MC) [64] para esse tipo de cálculo. Nessas simulações utilizam-se dinâmicas estocásticas para evoluir temporalmente um sistema. Neste caso produzimos uma cadeia markoviana de configurações na variável temporal não física τ dita de Monte Carlo. O algoritmo empregado deverá amostrar configurações estatisticamente distribuídas de acordo com um dado ensemble ao longo da evolução temporal da simulação. Assim, assumindo ergodicidade, substituem-se médias térmicas $\langle \dots \rangle_T$ de ensemble por médias temporais $\langle \dots \rangle_\tau$ nas simulação de MC.

Na iminência de uma transição de fase, a produção de configurações independentes é afetada fortemente pelo efeito do frenamento crítico (do inglês, *critical slowing down*: CSD) [63]. Este efeito está associado à existência de comprimentos de correlação divergentes em transições de fase de segunda ordem, e à presença de enormes barreiras interfaciais na energia livre canônica em transições de primeira ordem. Isto produz divergências nos tempos de autocorrelação entre configurações sucessivas. Para contornar o problema, sofisticados algoritmos de atualização globais devem ser empregados sempre que possível.

Outra alternativa é a utilização de ensembles generalizados em que o peso canônico usual de Boltzmann-Gibbs $e^{-\beta E}$, típico do ensemble canônico, é substituído por outros, com a finalidade de diminuir a influência numérica das barreiras de energia livre [61, 62, 77].

Iniciamos o capítulo revisando brevemente alguns tipos de simulações, o conceito de cadeias de Markov é então introduzido e utilizado na formulação do algoritmo de Metropolis. Discutimos em seguida como estimar erros numéricos e como o efeito do CSD pode afetá-los. Estas observações orientaram historicamente o desenvolvimento de algoritmos mais eficientes, chamados de microcanônicos, imunes ao efeito de CSD. Finalizamos este capítulo mostrando como adaptar os métodos de Monte Carlo usuais, conjuntamente com o ensemble multicanônico, para determinar densidades de estados.

3.1 Métodos de Monte Carlo markovianos

O cálculo de integrais multidimensionais é o objetivo de elaboradas técnicas numéricas. Uma abordagem bastante eficiente para estimá-las é o uso de algoritmos estatísticos, chamados de métodos de Monte Carlo. Sua implementação na presença de medidas de integração simples e homogêneas como é caso do exemplo a seguir,

$$I = \int_a^b \frac{f(x)}{b-a} dx, \quad (3.1)$$

se faz sorteando M valores aleatórios e independentes $x_i \in [a, b]$, segundo a medida (peso)

$$w(x) \equiv \frac{dx}{b-a}. \quad (3.2)$$

Nesta situação, a integral será bem aproximada pela média da função aleatória $f(x_i)$

$$I \cong \bar{f} \equiv \frac{1}{M} \sum_{i=1}^M f(x_i), \quad (3.3)$$

cujo erro numérico é dado pelo *desvio padrão* ($\sigma_{\bar{f}}$)

$$\sigma_{\bar{f}} = \sqrt{\frac{\langle [f(x_i) - \langle f(x_i) \rangle]^2 \rangle}{M-1}} \quad (3.4)$$

que tende a zero no limite $M \rightarrow \infty$. Este método é denominado Monte Carlo estático pois os sorteios ocorrem segundo a distribuição na Eq. (3.2) sem a necessidade de obedecer a qualquer mecanismo dinâmico.

A estimativa de valores esperados de observáveis mecânico-estatísticos $\langle O \rangle_\beta$ implica no cálculo de expressões como

$$\langle O \rangle_\beta = \frac{\int \prod_{i=1}^{i=N} d_i \vec{q} d_i \vec{p} e^{-\beta H_N(p,q)} O(p,q)}{\int \prod_{i=1}^{i=N} d_i \vec{q} d_i \vec{p} e^{-\beta H_N(p,q)}}, \quad (3.5)$$

em que um sorteio segundo a medida de integração $d\vec{x} = \prod_{i=1}^{i=N} d_i \vec{q} d_i \vec{p} e^{-\beta H_N(p,q)}$ torna-se inviável de acordo com os métodos estáticos. Para contornar este problema necessita-se de um processo estocástico¹ cuja dinâmica de equilíbrio reproduza a distribuição almejada. Esta técnica chama-se método dinâmico de Monte Carlo [63].

As dinâmicas estocásticas empregadas nos métodos dinâmicos de amostragem não guardam relação com a dinâmica física real do sistema investigado e por isso são denominadas

¹Denomina-se processo estocástico uma sequência de variáveis aleatórias, ou série temporal, cujos elementos são indexáveis em determinado intervalo temporal t .

simulações. Nesse contexto a utilização de *cadeias de Markov*, evoluindo sob algoritmos ergódicos², constitui a principal classe de métodos de Monte Carlo markovianos³ capazes de implementar estimativas como na Eq. (3.5).

Cadeias de Markov, por sua vez, são sequências X_1, X_2, X_3, \dots de variáveis aleatórias em um espaço de estados $\{X_n\}$, onde X_n é o estado deste processo estocástico no instante de tempo n . Se a distribuição de probabilidade condicional de X_{n+1} nos estados passados for uma função apenas de X_n , teremos

$$\Pr(X_{n+1} = x | X_0, X_1, X_2, \dots, X_n) = \Pr(X_{n+1} = x | X_n), \quad (3.6)$$

onde x é um estado do processo. Esta propriedade é dita markoviana⁴ e permite que a cada instante de tempo n caracterizemos a cadeia por uma matriz P que fornece probabilidades (estacionárias) de transição entre os estados i e j ,

$$P_{ij} = \Pr(X_{n+1} = j | X_n = i). \quad (3.7)$$

Como veremos a seguir, desenhar eficientes algoritmos estocásticos que permitam construir a matriz na Eq. (3.7) para pesos de integração específicos é uma arte computacional.

3.1.1 O algoritmo de Metropolis

O algoritmo que Metropolis e colaboradores desenvolveram [64] permite gerar distribuições de equilíbrio canônicas. Neste caso, a probabilidade de ocorrência $P(E)$ de uma energia E é dada por $P(E) \propto \Omega(E) e^{-\beta E}$, em que $\Omega(E)$ é a densidade de estados do sistema e $1/\beta$ é a temperatura do reservatório térmico. A implementação deste algoritmo necessita de sorteios aleatórios (e ergódicos) de configurações microscópicas do sistema. Também é preciso que se consiga avaliar as alterações energéticas surtidas por aquelas mudanças configuracionais aleatórias. O algoritmo garante então que, partindo-se de quaisquer condições iniciais, haverá convergência da distribuição de probabilidades para uma situação canônica de equilíbrio, ou seja $P(E) \rightarrow P_{can}(E) \equiv \Omega(E) e^{-\beta E}$.

Para ilustrá-lo considere um sistema hamiltonio $H = H(\{s_i\})$ dependente das configurações $\{s_i\}$. Implementamos o algoritmo de Metropolis para criar configurações $\{s_i\}$ de equilíbrio seguindo os passos:

1. Seja uma dada configuração inicial $\{s_1, \dots, s_N\}$. Seleccionamos um destes componentes, digamos s_i e proponhamos uma alteração aleatória nesta configuração. Ou seja, tomemos por exemplo $s_i \rightarrow s'_i$ e mantenhamos todos os demais elementos intactos.
2. Computemos a diferença de energias associadas à alteração na configuração $s_i \rightarrow s'_i$, ou seja: $\Delta H = H(s'_i) - H(s_i)$.

²Que permite a visitação de todos os estados acessíveis ao sistema.

³Do inglês: *Markovian Monte Carlo methods*

⁴Referindo-se aos processos estocásticos cuja matriz de transição, responsável por sua evolução temporal (de MC), implicam *dependência do futuro com o passado somente pelo passo presente* [78].

3. Calculemos a probabilidade de aceitação da nova configuração s'_i , utilizando a probabilidade de transição de Metropolis $P_{Metropolis}(s_i \rightarrow s'_i)$ dada por

$$P_{Metropolis}(s_i \rightarrow s'_i) = \begin{cases} 1 & \text{se } \Delta H < 0 \\ e^{-\beta \Delta H} & \text{se } \Delta H \geq 0. \end{cases} \quad (3.8)$$

4. Sorteamos um número r a partir de uma distribuição uniforme: $r \in [0, 1]$.
5. Se $r < P_{Metropolis}(s_i \rightarrow s'_i)$ substituímos s_i pela nova proposta s'_i , caso contrário, mantemos a antiga configuração s_i inalterada.

O sistema evolui com a aplicação dos 5 passos acima a cada um de seus componentes. Após certo número de atualizações configuracionais necessárias para atingir o equilíbrio, i.e., convergência para $P_{can}(E) \equiv \Omega(E) e^{-\beta E}$, podemos calcular médias como descritas na Eq. (3.5) com cada nova configuração $\{s_i\}$. Isto nos leva às médias estatísticas $\langle \mathcal{O}[\{s_i\}] \rangle_{\tau_{MC}}$ de funcionais $\mathcal{O}(\{s_i\})$ que são obtidas com M configurações $\{s_i\}$ desta série temporal e estimam as médias térmicas $\langle \mathcal{O} \rangle_{\beta}$.

3.1.2 Erros numéricos

Mencionamos na seção anterior que erros estatísticos em simulações de Monte Carlo são descritos pelo desvio padrão. Contudo, não levamos em conta possíveis correlações entre as sucessivas configurações mensuradas. Para fazê-lo teríamos de calcular o *tempo de autocorrelação de Monte Carlo*⁵, um parâmetro também associado à eficiência algorítmica.

A prescrição para este cálculo considera a série temporal de medidas de um observável físico \mathcal{O} num intervalo temporal t e a *autocorrelação normalizada* $\rho_f(k) = \frac{C_f(k)}{C_f(0)}$ [63, 78]. A autocorrelação é escrita em termos das correlações $C_f(k)$ e quantifica possíveis dependências estatísticas

$$\rho_{f\mathcal{O}}(k) \equiv \frac{C_{f\mathcal{O}}(k)}{C_{f\mathcal{O}}(0)} = \frac{\langle \mathcal{O}_i \mathcal{O}_{i+k} \rangle - \langle \mathcal{O}_i \rangle^2}{\langle \mathcal{O}_i^2 \rangle - \langle \mathcal{O}_i \rangle^2}. \quad (3.9)$$

Tipicamente a função de autocorrelação normalizada $\rho_f(t)$ decai exponencialmente ($\rho_f(t) \sim e^{-|t|/\tau_{exp,\mathcal{O}}}$) para grandes valores de t . Isto nos leva a definir o *tempo de autocorrelação exponencial* para a série do observável \mathcal{O} como

$$\tau_{exp,\mathcal{O}} = \limsup_{t \rightarrow \infty} \frac{t}{-\log |\rho_{f\mathcal{O}}(t)|}. \quad (3.10)$$

Ou seja, $\tau_{exp,\mathcal{O}}$ é uma escala de relaxação temporal correspondente ao modo mais lento de decaimento da correlação das amostras.

⁵Que é o intervalo do tempo computacional durante a produção de sucessivas amostras correlacionadas ou estatisticamente dependentes.

Por outro lado, é interessante calcular o *tempo integrado de autocorrelação*, que recai em Eq. (3.10) no limite $\tau \gg 1$

$$\tau_{int,\mathcal{O}} = \frac{1}{2} + \sum_{t=1}^{\infty} \rho_{f\mathcal{O}}(t). \quad (3.11)$$

Sokal demonstrou rigorosamente [78], sob a condição $N \gg \tau$, que na presença de correlações a Eq. (3.4) converte-se em

$$\begin{aligned} \sigma_f^2 &= \frac{1}{N^2} \sum_{r,s=1}^N C_f(r-s) \\ &= \frac{1}{N} \sum_{t=-(N-1)}^{N-1} \left(1 - \frac{|t|}{N}\right) C_f(r-s) \\ &\approx \frac{1}{N} (2\tau_{int,f}) C_f(0). \end{aligned} \quad (3.12)$$

Portanto, surge um termo de correção multiplicativo $\sqrt{2\tau_{int}}$ no erro padrão.

Adicionalmente, os tempos de correlação permitem caracterizar o fenômeno do frenamento crítico que desacelera as simulações de Monte Carlo ao redor de transições de fase. Por exemplo, foi demonstrado [78] que para um sistema finito com extensão linear L , nas vizinhanças de uma transição de fase contínua, o tempo integrado de autocorrelação de \mathcal{O} comporta-se como

$$\tau_{int,\mathcal{O}} \propto L^z. \quad (3.13)$$

Este expoente $z (\geq 1)$ é dito dinâmico e quantifica o desempenho dos algoritmos de Monte Carlo empregados. A procura por algoritmos mais eficientes, i.e. com menores valores de z , impulsionou o desenvolvimento das simulações microcanônicas apresentadas a seguir.

3.2 Simulações microcanônicas

Uma solução eficiente para contornar o efeito do CSD emprega ensembles generalizados [50, 54, 56, 61, 62], que permitem suavizar as barreiras na energia livre canônicas. A implementação computacional destes métodos é direta pois utilizam versões do algoritmo de Metropolis. Um subproduto desta abordagem é que ela pode produzir estimativas diretas da entropia microcanônica Eq. (2.1). Isto porque a relação Eq. (2.16) pode ser aplicada, após termos determinado os parâmetros multicanônicos,

$$S_{micro}(E) \equiv \ln(W) \cong b(E)E - a(E). \quad (3.14)$$

Podemos considerar as simulações multicanônicas consistindo de duas etapas distintas

1. *Determinação dos pesos multicanônicos*: nesta etapa uma prescrição para calcular os pesos multicanônicos é empregada, o que se faz geralmente por meio de *relações de recorrência*. Esta etapa será apresentada em detalhes a seguir e constitui as simulações microcanônicas propriamente ditas.
2. *Simulação multicanônica produtiva*: nessa etapa uma longa cadeia de Markov é produzida por um algoritmo de Monte Carlo dinâmico. Em geral, utiliza-se o algoritmo de Metropolis com peso multicanônico apresentado na Eq. (2.13), pois já teremos os seus parâmetros determinados na etapa anterior. A evolução de Metropolis ocorre como descrevemos, exceto pelo fato de que a probabilidade de transição dada pela Eq. (3.15) é modificada pela substituição $w_{canônica} \rightarrow w_{muca}$, o que resulta

$$P_{Metropolis-MUCA}(s_{velho} \rightarrow s_{novo}) = \begin{cases} 1 & \text{se } \Delta H < 0 \\ \frac{w_{muca}[H(s_{novo})]}{w_{muca}[H(s_{velho})]} & \text{se } \Delta H \geq 0. \end{cases} \quad (3.15)$$

3.2.1 Relações de recorrência

A determinação dos pesos multicanônicos $b(E)$ e $a(E)$ é normalmente feita por meio de uma série de simulações consecutivas, análogas à fase produtiva, em que controem-se histogramas das energias visitadas. Estes resultados são paulatinamente combinados via relações de recorrência, com as quais atualizam-se os pesos multicanônicos a cada simulação. Este processo é repetido diversas vezes para que ao final tenhamos a convergência de $w_{muca}(E) \rightarrow \frac{1}{\Omega(E)}$. Ou seja, a densidade de estados microcanônica seja computada com alta precisão.

Construiremos a seguir estas relações de recorrência. Consideremos por simplicidade os pesos iniciais

$$w_{muca}^0(E_k) = 1, \quad \text{para todo } k. \quad (3.16)$$

Esta condição descreve um sistema físico colocado a uma temperatura infinita pois $1/T = b(E) = 0$. Em seguida, uma simulação de Monte Carlo é executada usando-se inicialmente w_{muca}^0 como peso estatístico na Eq. (3.15). Durante a evolução computacional as energias E_k são mensuradas e geramos o histograma $H_{muca}^0(E)$.

Considerando que a entropia microcanônica é bem aproximada pelos parâmetros multicanônicos Eq. (3.14), empregamos a relação Eq. (2.3), isto é $\beta = (\partial S / \partial E)$, para definir

$$b(E) = \frac{\partial S(E)}{\partial E}, \quad (3.17)$$

ou, aplicando diferenças finitas

$$b(E) = \frac{S(E + \epsilon) - S(E)}{\epsilon}. \quad (3.18)$$

Portanto, partindo da definição de $S(E)$ apresentada na Eq.(3.14) podemos calcular a diferença

$$S(E) - S(E - \epsilon) = b(E)E - [b(E - \epsilon)](E - \epsilon) - a(E) + a(E - \epsilon). \quad (3.19)$$

Este resultado é simplificado com a substituição da identidade $S(E) - S(E - \epsilon) = \epsilon b(E - \epsilon)$, i.e., da Eq. (3.18) na expressão anterior originando uma relação entre os parâmetros multicanônicos

$$a(E - \epsilon) = a(E) + [b(E - \epsilon) - b(E)]E, \quad (3.20)$$

em que os $a(E)$ e $b(E)$ provém da mesma $n - \text{ésima}$ simulação.

Temos ainda que estabelecer uma relação de recorrência para os parâmetros multicanônicos $b^{n+1}(E)$ e $b^n(E)$, originários de simulações sucessivas. A implementação mais interessante [61, 62] usa os estimadores $\hat{H}_{muca}^n(E)$ para os histogramas de energias $H_{muca}^n(E)$

$$\hat{H}_{muca}^n(E) = \max[h_0, H_{muca}^n(E)]. \quad (3.21)$$

O uso da constante $h_0 \in (0, 1)$ impede o aparecimento de entradas nulas.

Definimos então

$$w_{muca}^{n+1}(E_k) = e^{-S^{n+1}(E)} \doteq c \frac{w_{muca}^n(E_k)}{\hat{H}_{muca}^n(E)}, \quad (3.22)$$

em que a constante c é introduzida para que $S^{n+1}(E)$ seja um estimador da entropia microcanônica

$$S^{n+1}(E) = -\ln(c) + S^n(E) + \ln[\hat{H}_{muca}^n(E)]. \quad (3.23)$$

Finalmente, se substituirmos a Eq. (3.23) na equação Eq. (3.18), encontramos a relação de recorrência para os $b^{n+1}(E)$ em termos dos $b^n(E)$ e dos histogramas das energias,

$$b^{n+1}(E) = b^n(E) + \left[\ln \hat{H}_{muca}^n(E + \epsilon) - \ln \hat{H}_{muca}^n(E) \right] / \epsilon. \quad (3.24)$$

3.2.2 Implementação alternativa

Berg e colaboradores [61, 62] mostraram que monitorando-se o comportamento das variâncias $\sigma^2[b^{n+1}(E)]$ dos parâmetros $b^n(E)$ pode-se acelerar as simulações MUCA. Para isso, ele introduziu um novo peso estatístico expresso como

$$g_0^n(E) = \frac{c}{\sigma^2[b^{n+1}(E)]} \equiv \frac{\hat{H}_{muca}^n(E + \epsilon) \hat{H}_{muca}^n(E)}{\hat{H}_{muca}^n(E + \epsilon) + \hat{H}_{muca}^n(E)}, \quad (3.25)$$

onde c é uma constante arbitrária. A seguinte relação de recorrência é obedecida

$$g^{n+1}(E) = g^n(E) + g_0^n(E), \quad g^0(E) = 0. \quad (3.26)$$

Mediante o emprego dos g' s nas simulações MUCA agiliza-se o aplainamento dos histogramas $H_{muca}^n(E)$, o que acelera e garante uma eficiente convergência $S_{muca} \rightarrow S_{micro}$. Para implementá-lo, basta substituirmos a Eq. (3.24) pela seguinte relação de recorrência

$$b^{n+1}(E) = b^n(E) + \hat{g}_0^n \left[\ln \hat{H}_{muca}^n(E + \epsilon) - \ln \hat{H}_{muca}^n(E) \right] / \epsilon,$$

onde os \hat{g}_0^n são pesos normalizados,

$$\hat{g}_0^n(E) \doteq \frac{g_0^n(E)}{g^n(E) + g_0^n(E)}. \quad (3.27)$$

Capítulo 4

Proteínas e o modelo AB

We may, I believe, anticipate that the chemist of the future who is interested in the structure of proteins, nucleic acids, polysaccharides, and other complex substances with high molecular weight will come to rely upon a new structural chemistry, involving precise geometrical relationships among the atoms in the molecules and the rigorous application of the new structural principles, and that great progress will be made, through this technique, in the attack, by chemical methods, on the problems of biology and medicine ...

[L. PAULING, Nobel Lecture, 11/12/1954]

Apresentamos neste capítulo uma breve revisão da bioquímica de heteropolímeros não ramificados provindos da condensação de aminoácidos, as proteínas [24]. Em seguida descrevemos sua composição química, estrutura geométrica e funcionalidades [23]. Abordamos também o fenômeno de enovelamento protéico pelo qual proteínas adquirem sua conformação espacial biologicamente funcional, dita nativa. Considerando que a formação de agregados protéicos é induzida geralmente por falhas no enovelamento protéico, o que ocasiona proteinopatias como os males de Parkinson [35], Alzheimer [36], Huntington [37] e a Diabetes II [84], revemos também esse tópico.

Neste contexto, enfocamos as proteínas chamadas Príons (*Proteinaceous and Infectious Particle*) [41] que, apesar de exercerem importantes funções biológicas, podem eventualmente deformarem-se em padrões patológicos autoreplicantes. Um fato notável sobre doenças priônicas, como os males de Creutzfeld-Jacob e a encefalopatia espongiforme [38], é a habilidade de saltarem barreiras interespecíficas sem sofrer mutações.

Por fim, apresentamos a metodologia *in silico* para o estudo protéico com a intenção de contemplar a modelagem teórica de proteínas. Não empregamos neste trabalho complexos campos de força interatômicos do tipo "all-atoms" [27] e sim suas versões efetivas ou "coarse-grained" [71]. Em particular, adotamos aqui o chamado modelo AB [68, 69] por sua simplicidade, generalidade e eficácia computacional.

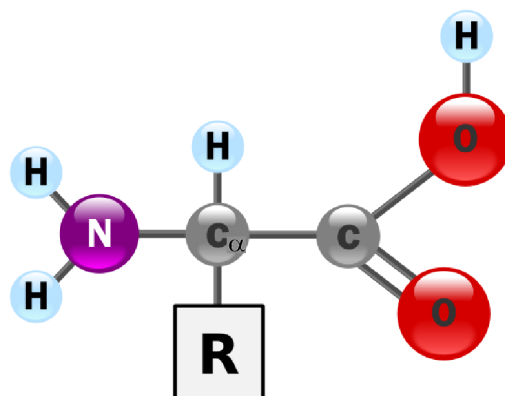


Figura 4.1: Estrutura geral de um aminoácido, com o carbono alfa C_α destacado ao centro, o grupo amina (NH_2) à esquerda, o carboxila (CO_2H) à direita e a cadeia lateral R abaixo.

Aminoácidos

Aminoácidos são compostos de carbono, hidrogênio, oxigênio, nitrogênio e às vezes, enxofre [23]. Sua estrutura, Figura (4.1), envolve um grupo amina (NH_2) e um grupo carboxila (CO_2H), ambos ligados a um carbono central, dito carbono alfa (C_α). O carbono alfa também está ligado a um hidrogênio e a uma cadeia lateral, que é representada pela letra R e identifica univocamente o aminoácido.

A fórmula molecular bidimensional esclarece parcialmente a estrutura dos aminoácidos, uma vez que suas mais importantes propriedades são naturalmente estereoquímicas ou tri-dimensionais. Aminoácidos possuem atividade óptica característica [23] e são encontrados em organismos vivos apenas sua forma levógira [31, 25].

Atualmente, conhecemos centenas de tipos de aminoácidos artificiais, ditos "não padrão" [25]. Porém, a evolução biológica selecionou apenas 20 destes, Figura (4.2), chamados de primários. Dentre eles, somente 8 são ditos essenciais¹, ou seja, não são produzidos pelo organismo humano. Podemos ainda classificá-los [24] em:

- Aminoácidos *apolares*: exibem radicais de hidrocarbonetos apolares ou modificados, exceção à glicina. Constituem radicais **hidrofóbicos**: alanina, leucina, valina, isoleucina, prolina, fenilalanina, triptofano e metionina.
- Aminoácidos *polares* (neutros): tem radicais formadores de pontes de hidrogênio, são **hidrófilos**. São eles glicina, serina, treonina, cisteína, tirosina, asparagina e glutamina.
- Aminoácidos *carregados*: tem caráter **hidrófilo**, podem ser ácidos com radicais carboxílicos, como os ácidos aspártico e glutâmico, ou básicos com radicais amina: arginina, lisina e histidina.

¹A saber: fenilalanina, isoleucina, leucina, lisina, metionina, treonina, triptofano e valina

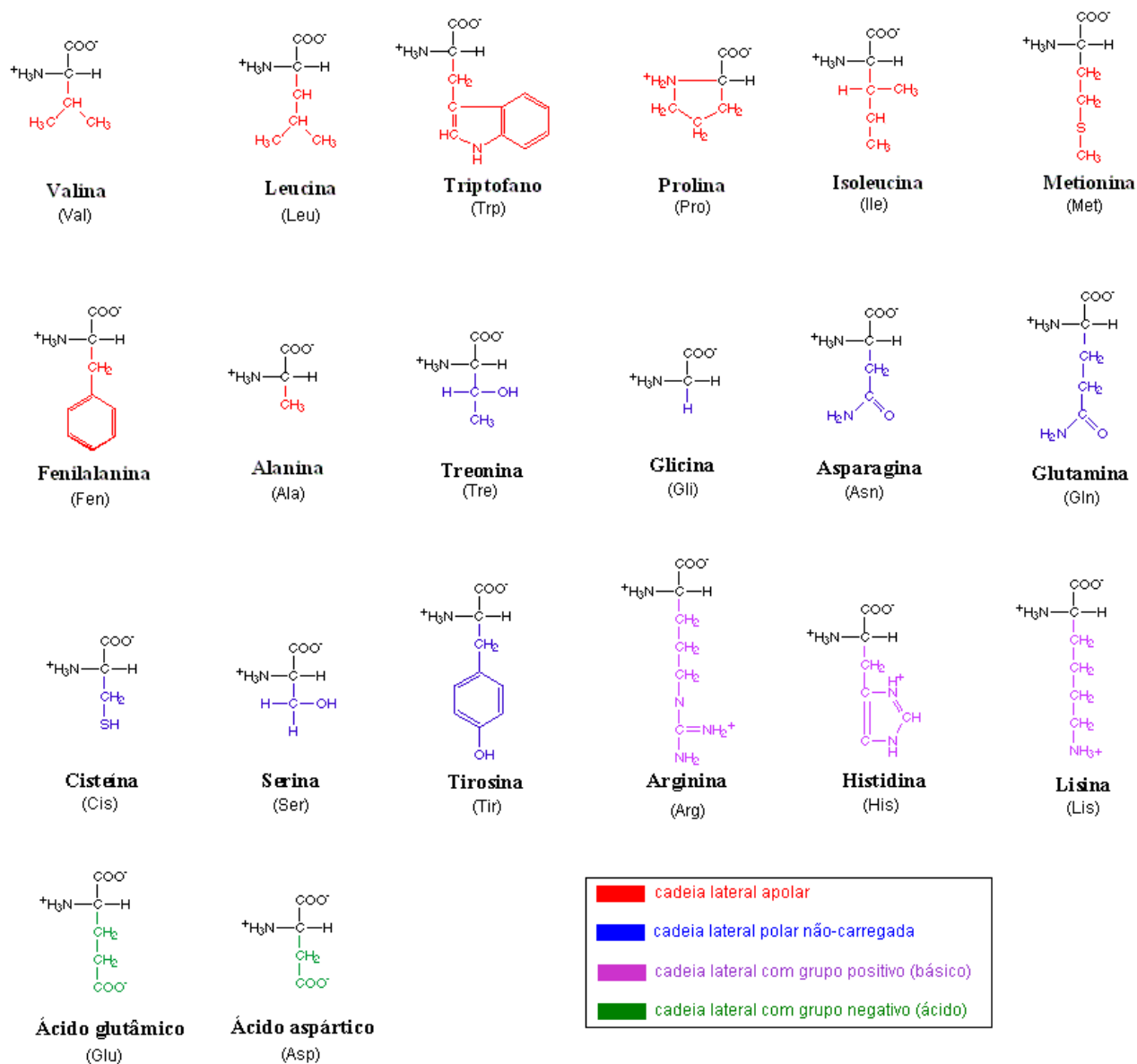


Figura 4.2: Estrutura química, nomenclatura, abreviação com código de 3 letras e classificação quanto a hidrofobicidade dos 20 aminoácidos primários.

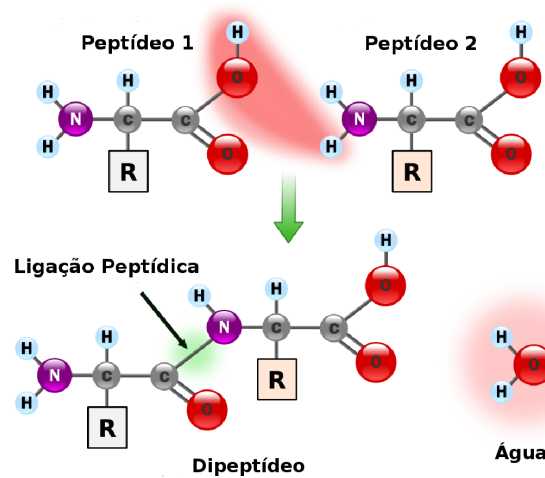


Figura 4.3: Formação de um dímero (dipeptídeo) pela condensação de aminoácidos.

4.1 Proteínas *in vitro* e *in vivo*: um panorama

Uma cadeia de aminoácidos é denominada peptídeo. Em geral denominamos pelo termo proteína [24] polipeptídeos enovelados e com função biológica definida, sendo compostas por centenas ou mesmo milhares de aminoácidos. Proteínas têm complexa estrutura e elevada massa molecular (de 5.000 a 1.000.000 u.m.a.). Elas podem ser sintetizadas por organismos vivos, o que se dá geralmente no citosol celular, pela condensação de aminoácidos que formam ligações peptídicas. Estas ligações tem caráter planar e estabelecem-se entre a amina de um aminoácido e o grupo carboxila do aminoácido seguinte com a eliminação de uma molécula de água, vide Figura (4.3). As proteínas são do ponto de vista estrutural [24] muito importantes em sistemas vivos [23]. Elas são classificadas em:

- *Globulares*: são o tipo mais comum de proteínas. Compactas, solúveis e dotadas de funções metabólicas, como a hemoglobina. São particularmente sensíveis a desnaturação ou destruição de sua estrutura terciária.
- *Fibrilares*: são geralmente cilíndricas, longas e filamentosas, insolúveis com tendência a agregarem-se, são robustas contra a desnaturação e possuem função estrutural somente em animais, como a queratina.
- *Membranosas*: são geralmente encontradas compondo a membrana celular, da qual somente separam-se por desnaturação ou sob ataque químico. Têm funções variadas como a de adesão celular, transportadoras, canais, receptoras e enzimáticas, como as glicoproteínas.

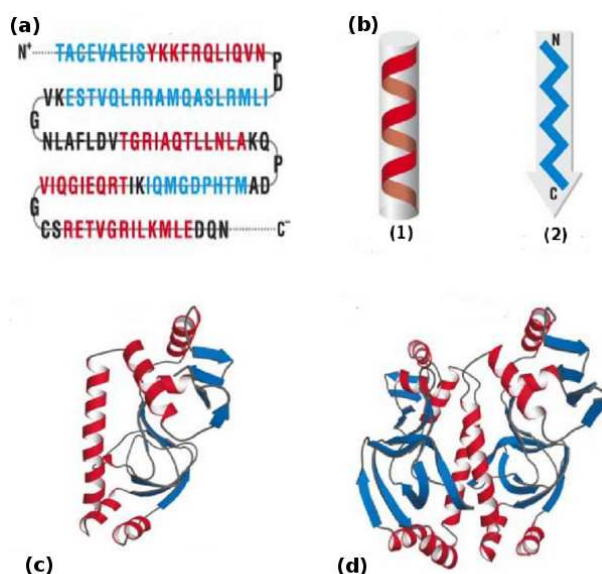


Figura 4.4: Hierarquia de estruturas em proteínas. (a) Primária: com abreviações de uma letra. (b) Secundária: (1) hélice alfa e (2) fita beta. (c) Terciária: associação de motivos enovelados, (d) Quaternária: união de cadeias peptídicas monoméricas.

4.1.1 Estrutura

O bioquímico dinamarquês Kai Linderstrom-Lang cunhou [24] os termos: estrutura primária, secundária e terciária para enfatizar a hierarquia estrutural em proteínas, Figura (4.4). Proteínas que tem somente uma cadeia são chamadas monoméricas, mas um grande número de proteínas são multimoméricas, com estruturas quaternárias como a hemoglobina. Apresentamos esta hierarquia estrutural a seguir.

(1) *Estrutura primária*, corresponde à seqüência de aminoácidos ao longo da cadeia polipeptídica. É o nível estrutural mais simples e mais importante, pois dele deriva todo o arranjo espacial da molécula. São específicas para cada proteína, sendo geralmente determinadas geneticamente. A estrutura primária da proteína resulta em uma longa cadeia de aminoácidos, semelhante a um colar de contas, com extremidades *amino terminal* e *terminal carboxílico*. Esta estrutura descreve somente a seqüência dos aminoácidos, ou resíduos, sem conexão com a orientação espacial da molécula. Quando duas proteínas têm suas estruturas primárias suficientemente semelhantes, elas são ditas homólogas.

(2) *Estrutura secundária*, é dada pelo arranjo espacial de aminoácidos próximos entre si na seqüência primária da proteína. Ocorre graças à possibilidade de rotação das ligações entre os carbonos dos aminoácidos e seus grupamentos amina e carboxila. O arranjo secundário de um polipeptídeo pode ocorrer de forma regular quando os ângulos das ligações entre carbonos alfa e seus ligantes são iguais e se repetem ao longo de um segmento

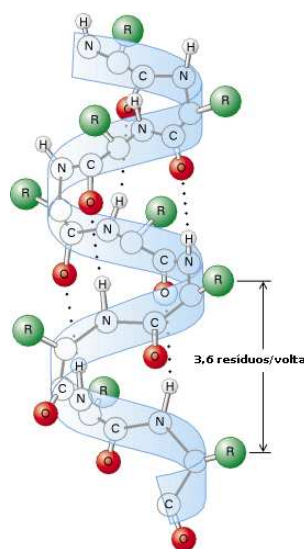


Figura 4.5: Estrutura secundária: arranjo espacial típico de uma *hélice alfa*, note as cadeias laterais externas ao eixo de simetria e a frequência espacial de 3,6 resíduos por volta [23].

da molécula. São dois os tipos principais de arranjos secundários regulares, os quais são classificados em:

- *Hélice alfa*: assemelha-se a uma escada em espiral, Figura (4.5). Essa é a mais típica estrutura secundária e foi descrita pioneiramente por Linus Pauling na Caltech em 1951 [23, 24]. Nela o esqueleto polipeptídico (*backbone*) está enrolado helicoidalmente ao longo de um eixo de simetria da molécula. Já as cadeias laterais (R) dos resíduos projetam-se para fora deste esqueleto. A estabilização se dá pela presença das ligações de hidrogênio entre os grupamentos NH e CO da cadeia principal com uma frequência média de 3,6 resíduos por volta. O sentido de giro de uma hélice alfa pode ser para a direita (hélice destrosa) ou para a esquerda (hélice sinistrosa). As hélices alfa encontradas em proteínas são destrosas [31]. Por causa da sua peculiar geometria as hélices alfa são polares e propiciam um momento de dipolo induzido. É esse arranjo que obriga a projeção das cadeias laterais para o exterior do *backbone*.
- *Folha beta*: é o segundo padrão estrutural mais comum em proteínas, Figura (4.6). Esse padrão foi proposto inicialmente por William Astbury em 1930 e posteriormente refinada por Linus Pauling e Robert Corey, 1951 [23, 24]. Nele as regiões vizinhas da cadeia polipeptídica associam-se por meio de ligações de hidrogênio, entre os grupos CO e NH, resultando em uma geometria achatada e rígida. Folhas beta são classificadas como paralelas (comuns), antiparalelas (relativamente comuns) ou mistas (incomuns), conforme as orientações relativas de sua cadeia principal (i.e., fita beta). Assim, dependendo da orientação relativa das fitas beta constituintes, podem surgir momentos de dipolo induzido entre as terminações carboxílicas (C) e amina (N).

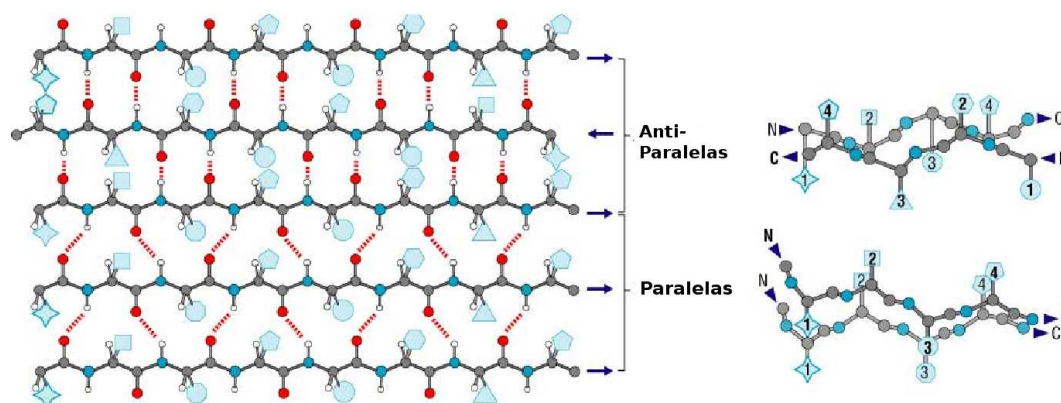


Figura 4.6: Estrutura secundária: conformações usuais de *folhas beta*, notam-se as fitas orientadas antiparalelas (superior) e paralelas (inferior), além das terminações amina (N) e carboxila (C).

Por exemplo, a folha pregueada é uma estrutura mantida por pontes de hidrogênio entre as unidades peptídicas. Neste caso, as ligações são estabelecidas entre cadeias polipeptídicas diferentes, distendidas e paralelas, ou entre segmentos distantes e distendidos em uma mesma cadeia. Vale notar ainda que existem associações de maior complexidade entre folhas beta, algumas delas têm sido associadas à formação de agregados protéicos e fibrilas. Esses agregados, por sua vez, são observados em diversas doenças humanas ditas amiloidoses (como no mal de Alzheimer), como veremos a seguir.

(3) *Estrutura terciária*, resulta do enovelamento ou dobramento da estrutura secundária, sendo estabilizada por pontes de hidrogênio e pontes dissulfeto. É esta estrutura que confere atividade biológica às proteínas. A estrutura terciária descreve a geometria tridimensional de uma cadeia resultante da interação de regiões, sejam elas estruturas regulares ou não. Neste nível organizacional pode ainda haver interações não-covalentes de segmentos distantes quanto à estrutura primária. Enquanto a estrutura secundária é determinada pelo relacionamento estrutural de curta distância, a terciária é caracterizada pelas interações de longa distância entre aminoácidos, denominadas *interações hidrofóbicas*, e pelas suas *interações eletrostáticas* [24].

Domínios e taxonomia de proteínas

A estrutura terciária pode ser interpretada como uma série de domínios agrupados. Nesse contexto definimos domínios como partes de um peptídeo que podem enovelar-se independentemente e são completamente funcionais. Baseados em argumentos de arranjos geométricos interconectados, M. Levitt e C. Chothia propuseram uma classificação taxonômica de proteínas [24], em que considera-se a presença majoritária dos domínios

- *Classes alfa*: tem seus núcleos formados exclusivamente por hélices alfa.

- *Classes beta*: tem seus núcleos formados exclusivamente de folhas beta, geralmente empacotadas umas contra as outras.
- *Classes alfa/beta*: apresentam combinações de motivos $\beta - \alpha - \beta$, tratando-se predominantemente de fitas beta paralelas circundadas por hélices alfa.
- *Classes alfa+beta*: são uma combinação discreta de hélices alfa e folhas beta.
- Outros: são geralmente domínios ricos em ligações dissulfídicas ou possuindo átomos metálicos. Assemelham-se a versões distorcidas de outras proteínas mais regulares.

(4) *Estrutura quaternária*, refere-se a composição das proteínas no mais alto nível de organização possível decorrente do fato que algumas proteínas podem ter duas ou mais cadeias polipeptídicas. Elas são guiadas e estabilizadas pelas mesmas interações presentes na estrutura terciária. Uma das principais representantes da estrutura quaternária é a hemoglobina, formada por quatro cadeias polipeptídicas, cujo mal agregacional ocasiona a anemia falciforme [24].

4.1.2 Funcionalidades

Proteínas desempenham diversas funções nos organismos vivos, dentre elas estão:

1. *Estrutural ou plástica*: dando rigidez, consistência e elasticidade ao tecido como o colágeno (das cartilagens) e a queratina (nos cabelos);
2. *Hormonal*: de ação específica sobre certos órgãos de um organismo como a insulina que retira a glicose em excesso do sangue;
3. *Defesa*: os anticorpos atuam no reconhecimento e neutralização de antígenos, enquanto peptídeos como o fibrinogênio e a trombina atuam na coagulação sanguínea.
4. *Energética*: fornecimento de energia a partir dos canais que compõem as proteínas;
5. *Enzimática*: catalisam reações bioquímicas específicas como as lipases que ajudam na hidrólise dos lipídeos;
6. *Condutoras de gases*: o transporte do oxigênio e parte do gás carbônico são realizados pela hemoglobina das hemácias.

4.1.3 Enovelamento protéico

O enovelamento protéico é um processo bioquímico fundamental, através dele a estrutura primária de uma proteína assume a sua configuração tridimensional biologicamente funcional, dita nativa. O processo contrário a este chama-se desnaturação, onde uma proteína é forçada a perder a sua configuração nativa, tornando-se uma cadeia de aminoácidos amorfa e não funcional. As proteínas desnaturadas podem perder a sua solubilidade e precipitar, tornando-se sólidos insolúveis.

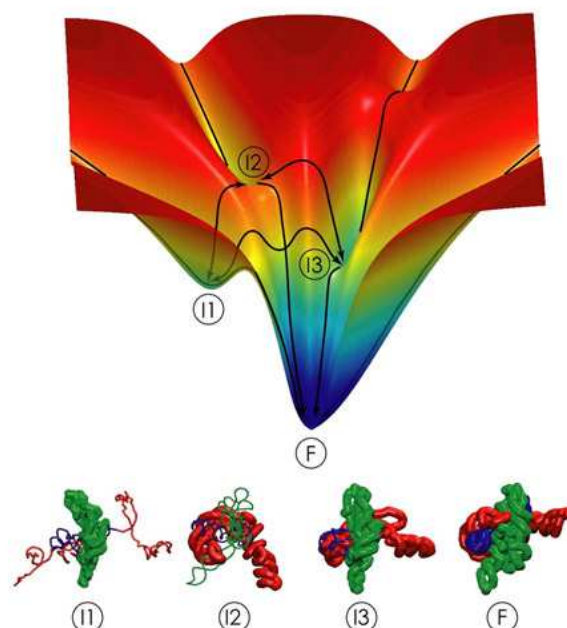


Figura 4.7: O perfil da energia livre (*energy landscape*) em formato de funil é um modelo aceito para a descrição do enovelamento protéico. Os estados 11, 12, 13 são intermediários enquanto "F" é a estrutura protéica nativa, ou de menor energia livre. As rotas preferenciais são representadas por setas.

A sequência específica de aminoácidos de uma proteína, ou estrutura primária, a condiciona a enovelar-se para adquirir sua configuração nativa. Muitas proteínas fazem-no espontaneamente durante, ou logo após a sua síntese no interior das células. Apesar destas macromoléculas aparentarem dobrar-se sozinhas, de fato seu enovelamento depende em grande medida das características da solução que as rodeia. Incluem-se aqui o tipo de solvente primário, que no interior das células são água e lípidos, a concentração dos sais, a temperatura, PH e o tipo das moléculas que a rodeiam. Por outro lado, há outras proteínas que somente enovelam-se com o auxílio de outras proteínas, as chaperonas [23].

Em um nível mais elementar, o enovelamento envolve inicialmente o estabelecimento de uma estrutura secundária, particularmente hélices alfa e folhas beta, formando a estrutura terciária. A formação de estruturas quaternárias aparenta envolver a montagem de subunidades previamente enoveladas [24]. Ao contrário das estruturas primárias, secundárias ou quaternárias, a estrutura terciária pode também envolver ligações covalentes na forma de pontes de dissulfeto entre dois resíduos de cisteína.

Um ponto essencial no enovelamento está no fato de que a sequência primária contém toda a informação necessária para atingir seu estado nativo: o enovelamento é um processo espontâneo (dogma de Anfinsen) [24]. A passagem ao estado enovelado é controlada por forças de Van der Waals [72] e por suas contribuições entrópicas à energia livre de Gibbs. Este aumento na entropia é conseguido movendo-se as seções hidrofóbicas das proteínas

para o interior e as hidrofílicas para o exterior, conferindo assim mais graus de liberdade às moléculas de água circundantes. Durante o processo de enovelamento, o número total de pontes de hidrogênio não muda apreciavelmente, para cada ponte de hidrogênio formada internamente à proteína outra com o meio aquoso é quebrada. Entretanto, o núcleo protéico pode ainda incorporar em si algumas moléculas de água.

Da tentativa de compreender o enovelamento protéico termodinamicamente surgiu o célebre paradoxo de Levinthal. Este paradoxo provém de um experimento mental realizado pelo físico e biólogo molecular C. Levinthal em 1961. Ele concluiu que o tempo de enovelamento de proteínas, considerando o perfil de suas energias livres como altamente frustradas (rugosas), seria várias ordens de magnitude superior ao observado [24]. Uma solução somente foi possível com a *teoria do Funil* [26], que sugere que perfis de energias livres protéicas são rugosas, mas afuniladas, possuindo rotas preferenciais para o enovelamento, Figura (4.7).

Observações experimentais evidenciam [32, 34] que peptídeos com mais de 100 resíduos apresentam geralmente os chamados estados intermediários durante o enovelamento, Figura (4.7). Contudo, não está plenamente elucidado se eles atuariam como auxiliares no processo ou se seriam antagonistas atuando como armadilhas locais. Embora nossa compreensão de como a sequência de aminoácidos produz detalhes do perfil de energia livre seja incipiente, suas características termodinâmicas estão certamente ligadas aos padrões primários de hidrofobicidade [26, 43, 44].

Acredita-se portanto, que o mecanismo de enovelamento envolveria um número relativamente pequeno de resíduos para a formação de um núcleo enovelante sobre o qual as demais estruturas condensar-se-iam rapidamente. Assim, uma vez que a topologia dos domínios envolvidos esteja corretamente estabelecida, a estrutura terciária formar-se-ia invariavelmente [32]. No caso contrário, quando tais interações fundamentais não se estabelecem, uma espécie de controle de qualidade entraria em ação e a proteína não poderia enovelar-se para sua estrutura nativa, sendo descartada [32].

4.2 Agregação protéica e proteinopatias

Certas doenças, conhecidas como proteinopatias, surgem na espécie humana como consequência de falhas durante o processo de enovelamento de algumas proteínas. As mais conhecidas são os males de Parkinson (Alfa-sinucleína) [35], Alzheimer (associado à proteína $A\beta$) [36], Huntington (Huntingtina) [37], Creutzfeld-Jacob (Príons) [38] e a Diabetes do tipo II (Amilina) [84], Figura (4.8). Invariavelmente essas falhas levam, por mecanismos ainda mal compreendidos², à formação de agregados fibrilares amilóides, ricos em fitas beta, altamente estruturados e resistentes à desnaturação [34]. Tais agregados acumulam-se em tecidos vivos, danificando-os e podendo levar o indivíduo à morte.

²Acredita-se, embora sem um consenso, que a formação de agregados amorfos após o desenovelamento parcial de proteínas tenha papel central neste processo. No caso das Príons, sabe-se que as estruturas em hélices alfa desnaturam-se para tornarem-se folhas beta, capazes de infectar outras Príons [34].

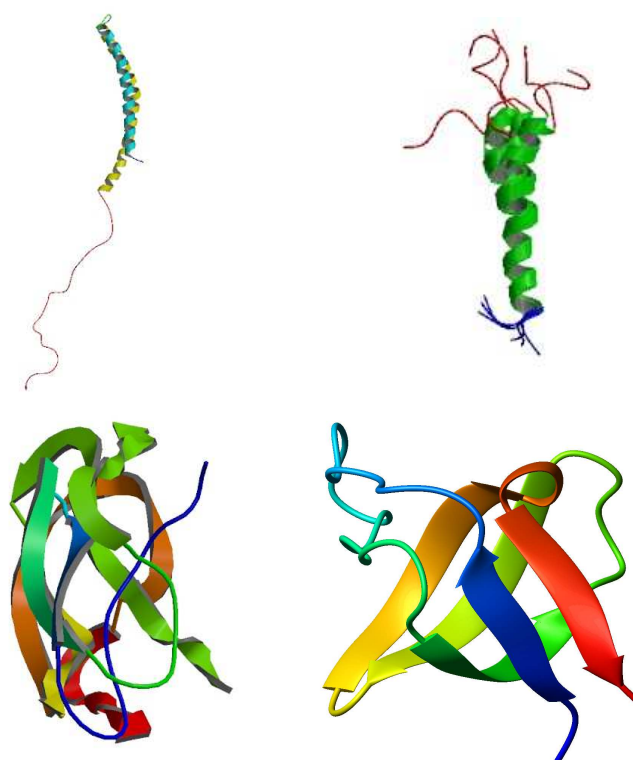


Figura 4.8: Proteínas envolvidas na formação de agregados. Superior à esquerda: *Alfa-Sinucleína*, associada ao mal de Parkinson. Superior à direita: *Amilina*, cuja formação de agregados ocorre na Diabetes do Tipo II. Inferior à esquerda: proteína precursora do *Amiloide-Beta* ($A\beta$), agregados de $A\beta$ ocasionam o mal de Alzheimer. Inferior à direita: domínio Src SH3, associado à transdução de sinais, não patológica. Adaptado do protein data bank (PDB).

Ainda no contexto da formação de fibrilas protéicas, observa-se também que outros peptídeos, como os domínios Src SH3 [34, 48], podem dar origem a agregados sem ocasionar proteinopatias. Curiosamente, contrário ao caso das Príons que são ricas em hélices alfa, os domínios Src SH3 possuem majoritariamente estruturas secundárias do tipo folhas beta. Essas constatações indicam que em certas condições há potencialidade agregacional em diversas proteínas, respeitadas as suas propensões específicas.

Assim, a formação de fibrilas amilóides parece ser um mecanismo genérico de agregação em que proteínas desestabilizadas são guiadas pelas interações hidrofóbicas. Entretanto, esse mecanismo requer algum ingrediente extra, necessário ao alinhamento ordenado de folhas beta para formar estruturas beta cruzadas como na Figura (4.9). Na inativação desse hipotético mecanismo observaríamos apenas agregados amorfos com elevado conteúdo de folhas beta mas sem a presença de hélices beta [73].

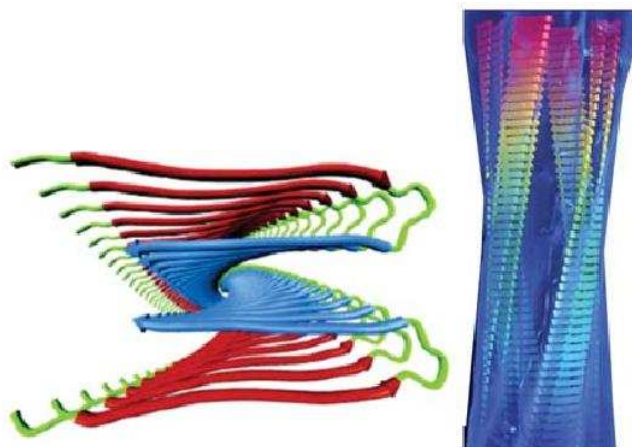


Figura 4.9: Modelo de fibrilas amilóides, ricas em formações de fitas beta como as hélices beta-cruzadas, que possuem alta resistência à desnaturação, [33].

Adicionalmente, a interconversão de agregados amorfos em fibrilas amilóides estruturadas [74] aponta para um mecanismo de rearranjo de fitas beta. Nesse caso, diversos estágios de desnaturações parciais, em diferentes níveis de intensidade, tomariam parte no processo. Ainda, a formação de fibrilas poderia ocorrer como um lento processo de nucleação, seguido de uma fase de montagem do agregado, em que a nucleação desempenharia papel essencial [33]. Ao contrário do que ocorre no processo de enovelamento, associado à interação terciária entre cadeias laterais, o fenômeno de agregação estaria ligado à interação entre as cadeias proteicas principais (*backbones*), enquanto as cadeias laterais atuariam de forma antagônica (prevenindo a agregação).

Existem também evidências crescentes [33, 34] de que cadeias desenoveladas não são aleatoriamente estruturadas. Ao contrário, elas poderiam ser entendidas como sistemas relativamente rígidos em razão das interações locais entre suas cadeias laterais e a principal. Também a estrutura topológica destas cadeias parece estar codificada pelas propensões estruturais de cada aminoácido na sequência primária. Por isto mesmo os estados altamente desenovelados ainda possuiriam uma topologia residual simples e sua geometria seria definida pelo caráter hidrofóbico - hidrofílico dos grupos locais de resíduos. Ainda que cadeias desenoveladas sejam apenas estados transientes, seu comportamento é crucial para o enovelamento e a agregação (como nucleadores) protéicos, conforme esquematizado na Figura (4.10).

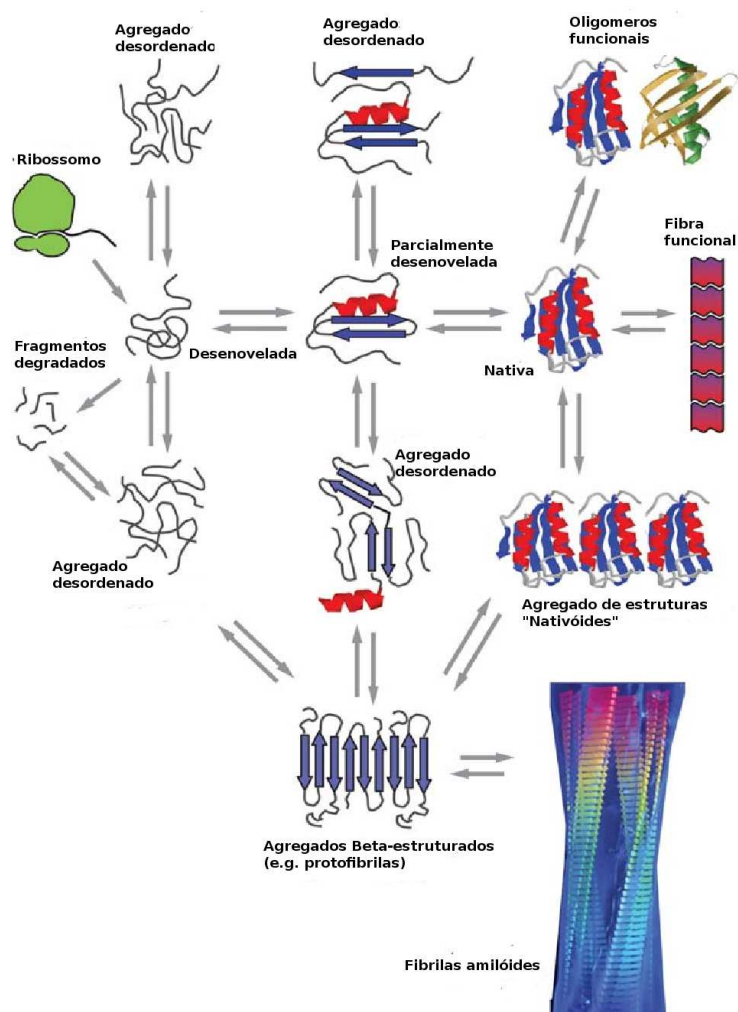


Figura 4.10: Representação esquemática dos possíveis estados conformacionais assumidos por cadeias peptídicas e suas interconversões usuais. Adaptado da Ref. [33].

Príons

A palavra príon vem do acrônimo de **PR**oteínacio, **I**nfeccioso e do afixo **ON** em analogia ao termo *virion*. Príons são agregados supramoleculares acelulares [39, 40, 41], compostos por proteínas (PrP^{Sc}) com capacidade de modificar outros indivíduos "sadios" (PrP^C), tornando-os cópias de si. Nesse processo, porções de estruturas do tipo hélice são renoveladas em folhas beta, processo este que induz profundas mudanças físico-químicas na proteína infectada [41].

Uma príon não possui ácido nucleico (DNA ou RNA), sendo o único tipo de protei-nopatia conhecido capaz de contagiar diretamente outros organismos e saltar barreiras interespecíficas. São conhecidos treze tipos de príons, dos quais três atacam fungos e dez afetam mamíferos [40]; dentre estes, sete têm por alvo a nossa espécie. As príons são causa-

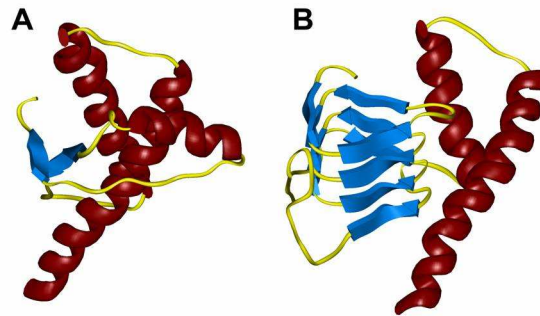


Figura 4.11: Configurações assumidas por Prions: a forma usual PrP^C representada por (A); e a patogênica PrP^{Sc} associada aos males de Creutzfeldt-Jakob e à Encefalopatia Espongiforme ilustrada em (B).

doras de várias doenças, como a encefalopatia espongiforme bovina, vulgarmente conhecida como "doença da vaca louca", e a doença de Creutzfeldt-Jakob (DCJ) [38].

Todas as doenças causadas por príons afetam a estrutura do cérebro ou dos tecidos neurais, não possuem cura e são sempre fatais. Normalmente as príons celulares PrP^C protegem os neurônios da morte e estão envolvidas no mecanismo de diferenciação neuronal [40]. Por isso inexistem defesas imunológicas naturais capazes de eliminar suas variantes patogênicas.

Enfermidades priônicas, seja em animais ou em humanos, diferenciam-se notadamente por suas variações no tempo de incubação, distribuição espacial das lesões e potencial infeccioso. Invariavelmente elas causam aglomerados tóxicos no cérebro, formando estruturas denominadas *placas amilóides*. Acredita-se que a compreensão da formação destas estruturas possa identificar mecanismos operacionais universais. Se este for o caso, ampliar-se-á inclusive o entendimento de outras amiloidoses neurodegenerativas como os males de Parkinson [35] e de Alzheimer [36]. O que poderá abrir caminho a uma grande classe de novas terapias.

No que concerne às peculiaridades das encefalopatias espongiformes priônicas, acredita-se atualmente [33, 34, 40, 41] que, diferentemente das demais proteinopatias, as PrP^{Sc} permaneçam infectantes mesmo quando parcialmente desnaturadas³. Esta visão encontra suporte em estudos envolvendo mutagênese, os quais evidenciam que modificações na estabilidade protéica das PrP^{Sc} não afetam diretamente sua tendência em produzir fibrilas [34]. Embora em PHs fisiológicos as PrP^C pareçam desenovelar-se mediante mecanismos de dois estágios, ou seja, sem a presença de estados intermediários, em PH ácido o mesmo processo revela estados intermediários ricos em estruturas beta. Conjectura-se que estes estados intermediários sejam os responsáveis termodinâmicos pela formação das PrP^{Sc} [34].

³Dai sua notável habilidade de infectar humanos a partir da ingestão de carne animal mesmo se cozida.

4.3 Proteínas *in silico*

Presenciamos um vertiginoso crescimento do poder computacional durante as últimas décadas. Este fenômeno tecnológico tem impulsionado o estudo numérico com técnicas *ab initio* de diversos sistemas complexos. Por isto as simulações de alto desempenho, efetuadas em supercomputadores [85], tem papel central e status de quase-experimentos em muitas áreas da física. Neste contexto o termo proteínas *in silico* aparece em contradistinção aos estudos bioquímicos tradicionais, baseados em experimentos *in vitro*.

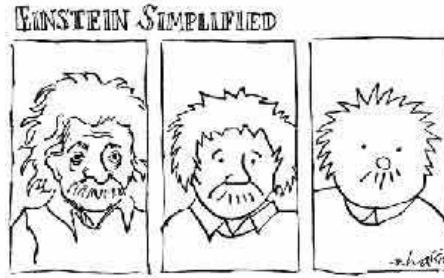
Inicialmente, devemos considerar a mais conveniente modelagem física das proteínas. Pode-se por exemplo partir de diferentes níveis de detalhamento estrutural, dependendo da tarefa a ser executada. No caso mais fundamental estão as descrições efetuadas em nível atômico por meio de modelos conhecidos como *all-atoms*. Nessas abordagens potenciais refinados [27, 28, 29, 30] são empregados para obter descrições peptídicas detalhadas. Entretanto, o alto custo computacional envolvido nesses estudos é um fator limitante, especialmente em se tratando de grandes proteínas. Uma alternativa para abrandar esta dificuldade vem do uso de modelos mínimos ou *coarse-grained* [67, 71]. Estes modelos não se baseiam em potenciais atômicos fundamentais, mas sim em interações efetivas estabelecidas entre representações das cadeias principais, laterais e do solvente.

É bem sabido que interações de longo alcance como as de caráter hidrofóbico-hidrofílico entre aminoácidos, modeladas por potenciais à la Lennard-Jones [72], são indispensáveis ao enovelamento e a formação de agregados. Portanto, é natural que elas sejam incluídas em quaisquer modelos mínimos, com variados graus de sofisticação.

Por outro lado, a correta formação de estruturas secundárias deve-se a interações locais (covalentes) estabilizadas por pontes de hidrogênio, um mecanismo cuja modelagem é delicada, sendo geralmente aproximada por potenciais do tipo mola + torção ou tubo flexível [65, 66, 71].

Portanto, o poder preditivo e a compatibilidade bioquímica das descrições implementadas por um ou outro modelo variam notavelmente, o que caracteriza a arte do modelamento matemático, já que muitas vezes é preciso substituímos a acurácia física de uma abordagem pela simplicidade computacional de outra.

4.3.1 O modelo AB *coarse-grained*



Things should be made as simple as possible, but not any simpler.
[A. EINSTEIN]

Para contornar obstáculos computacionais inerentes aos estudos de sistemas complexos, adotamos um modelo minimalista de proteínas proposto em [68, 69] e posteriormente generalizado para a aplicação em sistemas peptídicos (artificiais) interagentes [43, 44, 46, 65, 66, 70]. Este modelo é conhecido por AB e é formulado em um espaço contínuo (i.e., *off-lattice*). Nele interações de longo alcance entre aminoácidos são simplificadas notavelmente, empregando-se apenas descrições efetivas entre resíduos centrados nos carbonos alfa. Simplificam-se também as descrições das cadeias laterais, que são mapeadas segundo seu caráter hidrofóbico ou hidrofílico respectivamente em monômeros do tipo A ou do tipo B.

Uma vez que somente os carbonos alfa de cada resíduo são utilizados, formando assim uma cadeia principal ou *backbone*, sem cadeias laterais, a interação entre os resíduos fica reduzida à soma de dois termos hamiltonianos. O primeiro termo é um potencial torcional que tende a manter a cadeia distendida. Ao passo que o segundo termo corresponde ao potencial de Lennard-Jones, que descreve apropriadamente as interações hidrofóbicas relativas entre os resíduos [72].

Neste modelo a hamiltoniana de um sistema composto por uma única cadeia peptídica, dotada de N aminoácidos, é dada por

$$H_{1-prot.} = \frac{1}{4} \sum_{k=1}^{N-2} (1 - \cos \theta_k) + 4 \sum_{i=1}^{N-2} \sum_{j=i+2}^N \left(\frac{1}{r_{ij}^{12}} - \frac{C(\sigma_i, \sigma_j)}{r_{ij}^6} \right), \quad (4.1)$$

onde $0 \leq \theta_k \leq \pi$ é o ângulo diedral entre sucessivas ligações, enquanto r_{ij} denota a distância entre os i -ésimo e j -ésimo resíduos efetivos, vide esquema na Figura (4.12). O primeiro termo na Eq.(4.1) tende a distender a cadeia peptídica. Enquanto o segundo termo, que depende da distância entre os resíduos, introduz competição no sistema. Os coeficientes $C(\sigma_i, \sigma_j)$ são definidos segundo

$$C(\sigma_i, \sigma_j) = \begin{cases} +1.0 & \sigma_i, \sigma_j = A \\ +0.5 & \sigma_i, \sigma_j = B \\ -0.5 & \sigma_i \neq \sigma_j. \end{cases} \quad (4.2)$$

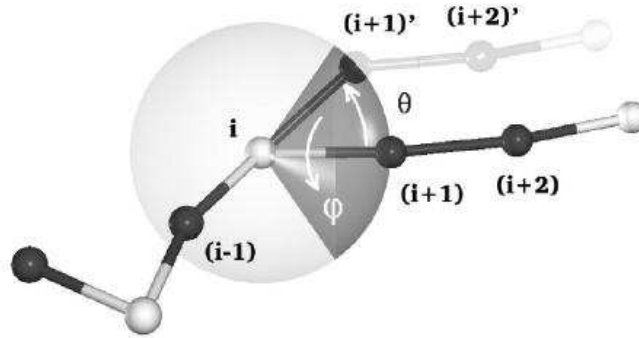


Figura 4.12: Esquema de um único peptídeo no modelo AB com N resíduos [46, 65, 87]. Note a representação da cadeia principal dos carbonos alfa (esferas), assim como seus $2N$ graus de liberdade de rotação (i.e. ângulos θ e φ).

Interações mais complexas poderiam ser facilmente introduzidas nessa abordagem *coarse-grained*, logicamente ao custo de maiores exigências computacionais. Por exemplo, poderíamos introduzir interações moleculares efetivas por pontes de hidrogênio utilizando modelos do tipo tubos flexíveis [44]. Ou mesmo, ainda seria possível otimizar a acurácia preditiva da formação de estruturas secundárias em proteínas reais com modificações no potencial torcional [67, 71]. Todas essas abordagens são desejáveis e válidas, constituindo entretanto objetivos a serem implementados apenas após exauridas as possibilidades investigativas do modelo AB original.

A simplicidade característica deste modelo é útil também para a análise computacional da formação de agregados. Para fazê-lo, uma generalização imediata do potencial na Eq. (4.1) é necessária de modo a permitir que um conjunto de polipeptídeos possa interagir de modo fisicamente plausível [43, 44, 46]. Consideremos então uma hamiltoniana composta pela soma das energias de cada uma das M proteínas individuais como na Eq. (4.1). Acrescentemos agora a ela um termo de interação intermolecular que envolva os pares de resíduos (i_μ, j_ν) situados em proteínas " μ " e " ν " distintas

$$H_{M\text{-prots.}} = \sum_{\mu=1}^{\mu=M} H_{1\text{-prot.}}^{\mu} + 4 \sum_{j_\mu > i_\nu + 1, i \neq j} \left(\frac{1}{r_{i_\mu j_\nu}^{12}} - \frac{C(\sigma_i, \sigma_j)}{r_{i_\mu j_\nu}^6} \right). \quad (4.3)$$

O modelo resultante apresenta interações competitivas entre partículas distantes, uma característica importante dos sistemas "pequenos" já observada em sistemas de spin [10]. Muitas vezes, a presença deste tipo de interação é sinônimo do aparecimento de quebra de ergodicidade e inequivalência de ensembles, revelando uma física bastante envolvente. Análises recentes com este modelo [43, 44, 46], mas empregando peptídeos artificiais (Fibonacci), relatam diversas peculiaridades da termodinâmica microcanônica apresentadas no Capítulo 2.

Capítulo 5

Resultados exatos: modelo de spin

*Wir bauen and Dir mit zitternden Händen
und wir türmen Atom auf Atom.
Aber wer kann Dich vollenden.
Du dom.*¹
[R. M. RILKE, Duino Elegies, 1923].

Como parte de nossos estudos desenvolvemos soluções exatas no ensemble gaussiano generalizado [83] para um sistema de spin com interações de longo alcance, o modelo Blume-Capel (BC) de alcance infinito [9, 15, 16, 79, 80]. A termodinâmica desse modelo é envolvente pois, embora exista seu limite termodinâmico, a física resultante é típica de um sistema "pequeno". Observamos aqui funções resposta, como calor específico e susceptibilidade magnética assumindo valores negativos [9, 10, 11, 12, 13]. Como consequência temos a inequivalência dos ensembles canônico e microcanônico [13, 17, 18, 19, 20, 21, 22, 59].

Também incorporamos neste capítulo conceitos apreendidos durante uma ampla revisão bibliográfica sobre ensembles generalizados [2, 6, 7, 17, 18, 19, 20, 21, 22, 49, 50, 51, 52, 53, 54, 55, 56, 58, 60, 61, 62, 75, 76, 82], pois suas interconexões e aplicações numéricas são essenciais a esta tese. Portanto, o estudo de soluções exatas apresentadas na referência [83] serve também de laboratório e introdução às técnicas da termoestatística microcanônica, eficazes na caracterização de transições de fase [2, 6, 7, 75, 76].

5.1 O modelo Blume-Capel no ensemble gaussiano estendido

O modelo Blume-Capel (BC) com interações de alcance infinito apresenta soluções analíticas nos ensembles canônico e microcanônico. Logo, seus diagramas de fases são

¹Com as mãos trêmulas te construímos; Átomo sobre átomo as tuas torres elevamos; Mas quem te poderá completar; Ó Catedral.

conhecidos nas duas abordagens [9, 10, 79, 80, 81]. Este modelo é dotado de riqueza fenomenológica evidenciada pela presença de transições de fase de primeira e segunda ordem, além de um ponto tricrítico [9]. Também a termodinâmica microcanônica do modelo exibe peculiaridades adicionais como: quebra de ergodicidade, entropia localmente não-côncava, calor específico e susceptibilidade magnética negativos e descontinuidades na temperatura como função da energia.

Propomos aqui uma resolução exata do modelo BC utilizando o ensemble gaussiano estendido, o qual interpola suavemente, de modo paramétrico (γ), as soluções nos ensembles canônico e microcanônico [83]. Tal abordagem permite-nos analisar o mecanismo gerador desta inequivalência, mostrando como estados canonicamente metaestáveis ou instáveis, embora microcanonicamente estáveis, são paulatinamente acessados pelo EGE conforme $\gamma \rightarrow \infty$.

Em adição, introduzimos uma nova forma de estudar pontos tricríticos na presença de inequivalências de ensembles. O conceito chave vem de exigir estabilidade termodinâmica da solução no EGE, impondo concavidade local à entropia generalizada (s_γ^*), como função do parâmetro γ , o que permite *quantificar* esta inequivalência. Por outro lado, nossa proposta também fornece um algoritmo para localizar pontos tricríticos, efetuando simulações com uma variação sistemática do parâmetro interpolante.

5.1.1 Um exemplo de sistemas "pequenos"

O modelo Blume-Capel é análogo ao modelo de Ising mas tendo spins $S_i = \{0, \pm 1\}$ e foi idealizado para descrever separações de fases em sistemas magnéticos [79]. Podemos também considerá-lo como um caso particular do modelo Blume-Emery-Griffiths [80], introduzido para explicar a coexistência de misturas de $He^3 - He^4$. A versão aqui estudada é a de alcance infinito, cuja hamiltoniana é

$$H(S) = -\frac{J}{2N} \left(\sum_{i=1}^N S_i \right)^2 + \Delta \sum_{i=1}^N S_i^2. \quad (5.1)$$

Interações de alcance infinito geralmente induzem não-extensividades que proíbem a tomada do limite termodinâmico [3]. Isto pode ser remediado introduzindo-se no acoplamento J uma dependência com o número de partículas (i.e. $J/2N$), o que restaura a extensividade mas não a aditividade da energia. Como as interações na Eq. (5.1) são de alcance infinito este sistema é considerado "pequeno", independentemente de N , e por isso tem peculiaridades termodinâmicas. Por exemplo, diferem seus pontos tricríticos canônico localizado em $(\Delta/J, T/J) = (0.46209812, 1/3)$ e microcanônico: $(\Delta/J, T/J) = (0.46240788, 0.33034383)$.

5.2 Solução exata no ensemble gaussiano estendido

Para resolvermos o modelo Blume-Capel no ensemble gaussiano estendido substituímos inicialmente a sua hamiltoniana, dada pela Eq. (5.1), na função de partição generalizada

deste ensemble, encontrada em Eq. (2.23). Deste modo obtemos a seguinte expressão formal

$$Z_\gamma(U, \alpha) = \sum_{\{S\}} e^{-\alpha H(S) - \gamma [H(S) - U]^2}, \quad (5.2)$$

com $\{S\}$ denotando soma sobre as configurações do sistema.

Na exponencial acima aparecem termos quadráticos em $H(S)$ que devem ser linearizados para permitir uma integração análoga ao caso canônico [79]. Podemos fazê-lo por meio de uma transformação de Hubbard-Stratonovich (HS)

$$e^{-bx^2/2} = \frac{1}{\sqrt{2\pi b}} \int_{-\infty}^{+\infty} dy e^{-y^2/2b - ixy}, \quad (5.3)$$

em que tomamos $b = 2\gamma$ e $x = H(S) - U$ e obtemos explicitamente

$$Z_\gamma(U, \alpha) = \sum_{\{S\}} \frac{1}{\sqrt{4\pi\gamma}} e^{-\alpha U} \int_{-\infty}^{+\infty} dy e^{-\frac{y^2}{4\gamma} - (iy+\alpha)(\Delta \sum_{i=1}^N S_i^2 - U)} e^{(iy+\alpha)(\sqrt{\frac{J}{2N}} \sum_{i=1}^N S_i)^2}. \quad (5.4)$$

A expressão acima ainda não pode ser somada diretamente. Para consegui-lo precisamos aplicar outra vez uma transformação de HS, mas escrita agora como

$$e^{ba^2} = \sqrt{\frac{b}{\pi}} \int_{-\infty}^{+\infty} dz e^{-bz^2 + 2baz}, \quad (5.5)$$

em que utilizamos $a = \sqrt{\frac{J}{2N}} \sum_{i=1}^N S_i$ e $b = iy + \alpha$, resultando em

$$\begin{aligned} Z_\gamma(U, \alpha) &= \frac{1}{\sqrt{4\pi\gamma}} \sum_{\{S\}} e^{-\alpha U} \int_{-\infty}^{+\infty} dy \left(\frac{iy+\alpha}{\pi}\right)^{1/2} e^{-\frac{y^2}{4\gamma} - (iy+\alpha)(\Delta \sum_{i=1}^N S_i^2 - U)} \\ &\quad \times \int_{-\infty}^{+\infty} dz e^{-(iy+\alpha)z^2 + 2(iy+\alpha)z \sqrt{\frac{J}{2N}} \sum_{i=1}^N S_i}. \end{aligned} \quad (5.6)$$

Observamos que o resultado acima simplifica-se se somarmos sobre os spins $S_i = \{0, \pm 1\}$. Então, efetuamos a soma em $\{S\}$ e aplicamos uma fatoração, reconhecendo a

presença de uma dupla expansão binomial, como mostrado pelas identidades a seguir

$$\begin{aligned}
 & \sum_{\{S\}} e^{-(iy+\alpha)\Delta \sum_i S_i^2 + 2(iy+\alpha)z \sqrt{\frac{J}{2N}} \sum_i S_i} \\
 &= \left[1 + e^{-(iy+\alpha)\Delta} \left(e^{2(iy+\alpha)z \sqrt{\frac{J}{2N}}} + e^{-2(iy+\alpha)z \sqrt{\frac{J}{2N}}} \right) \right]^N \\
 &= \sum_{N_0=0}^N \sum_{N_+=0}^{N-N_0} \frac{N!}{N_0!N_+!N_-!} e^{-(iy+\alpha)\Delta(N-N_0)} e^{-2(iy+\alpha)z \sqrt{\frac{J}{2N}}(N-N_0-N_+)} e^{2(iy+\alpha)z \sqrt{\frac{J}{2N}}N_+}.
 \end{aligned} \tag{5.7}$$

Substituindo agora o resultado da Eq. (5.7), que é a degenerescência dos estados do sistema, na função de partição Eq. (5.6) obtemos

$$\begin{aligned}
 Z_\gamma(U, \alpha) &= \frac{1}{\sqrt{4\pi\gamma}} e^{-\alpha U} \sum_{N_0=0}^N \sum_{N_+=0}^{N-N_0} \frac{N!}{N_0!N_+!N_-!} \\
 &\times \int_{-\infty}^{+\infty} dy \left(\frac{iy+\alpha}{\pi} \right)^{1/2} e^{-\frac{y^2}{4\gamma} - (iy+\alpha)(\Delta Q - U)} \int_{-\infty}^{+\infty} dz e^{-(iy+\alpha)z^2 + 2(iy+\alpha) \sqrt{\frac{J}{2N}} Mz},
 \end{aligned} \tag{5.8}$$

em que fizemos uso dos parâmetros de ordem do sistema $Q = \sum_{i=1}^N S_i^2 = N_+ + N_-$, $M = \sum_{i=1}^N S_i = N_+ - N_-$ e do vínculo sobre o número total de spins $N = N_+ + N_- + N_0$. Aqui, N_+ , N_- e N_0 denotam respectivamente o número de spins $+1$, -1 e 0 .

Finalmente, a Eq.(5.8) é integrada por meio de fórmulas gaussianas usuais e produz

$$Z_\gamma(U, \alpha) = \sum_{N_0=0}^N \sum_{N_+=0}^{N-N_0} \frac{N!}{N_0!N_+!N_-!} e^{-\alpha(\Delta Q - \frac{J}{2N}M^2) - \gamma(\Delta Q - U - \frac{J}{2N}M^2)^2}, \tag{5.9}$$

que simplifica-se ao escrevermos N_+ , N_- e N_0 em função de M e Q . Surge assim a solução exata:

$$Z_\gamma(U, \alpha) = \sum_{Q=0}^N \sum_{M=-Q}^{M=Q} \frac{N!}{(N-Q)! \left[\frac{1}{2}(Q+M) \right]! \left[\frac{1}{2}(Q-M) \right]!} e^{-\alpha(\Delta Q - \frac{JM^2}{2N}) - \gamma(\Delta Q - \frac{JM^2}{2N} - U)^2}. \tag{5.10}$$

5.3 Limites termodinâmicos: inequivalência de ensembles

Sabemos que o EGE possui uma conexão termodinâmica via potencial generalizado, dado pela Eq. (2.24), que permite-lhe recuperar tanto soluções canônicas quanto microcanônicas conforme os valores assumidos pelo parâmetro γ . Vale notar que ao efetuarmos o limite termodinâmico o parâmetro γ deve ser renormalizado, ou seja $\gamma \rightarrow \bar{\gamma} = N\gamma$, mas mantemos

a notação usual (i.e. $\bar{\gamma} = \gamma$) a fim de não sobrecarregá-la. Nesta seção pretendemos entender como a solução obtida anteriormente, do modelo BC no EGE, comporta-se nos limites: canônico ($\gamma = 0$) e microcanônico ($\gamma \rightarrow \infty$) que são inequivalentes neste modelo.

Iniciamos esta análise partindo da Eq. (5.10) que uma vez substituída na Eq. (2.24) permite expressar o seguinte potencial termodinâmico generalizado

$$\begin{aligned} \Phi_\gamma(\varepsilon, \alpha, m, q) = & \left[q \ln \left(\frac{\sqrt{q^2 - m^2}}{2(1-q)} \right) + \frac{m}{2} \ln \left(\frac{q+m}{q-m} \right) + \ln(1-q) \right] \\ & + \gamma \Delta^2 (q - Km^2 - \varepsilon)^2 + \alpha \Delta (q - Km^2). \end{aligned} \quad (5.11)$$

Note que utilizamos acima as seguintes quantidades específicas $\varepsilon = U/\Delta N$, $q = Q/N$, $m = M/N$, o acoplamento $K = J/2\Delta$, além de empregar aproximações de ponto de sela — i.e. $Z(\varepsilon, \alpha) \approx e^{-N\Phi(\varepsilon, \alpha, m)}$, vide Apêndice A — e de Stirling [9, 15, 16] no limite termodinâmico $N \rightarrow \infty$. Claramente, o termo entre os colchetes na Eq. (5.11) é o negativo da entropia microcanônica $s_{micro}(\varepsilon, m, q)$ obtida na referência [9].

Podemos calcular com a substituição do potencial Eq. (5.11) na equação Eq. (2.26) a entropia (específica) do EGE, o que resulta em $s_\gamma = s_{micro}$. A partir deste resultado e da definição na Eq. (2.19) encontramos a curva calórica do modelo BC neste ensemble

$$\alpha(\varepsilon) = \frac{1}{\Delta} \frac{\partial s_\gamma}{\partial \varepsilon}, \quad (5.12)$$

em que α é análoga ao inverso da temperatura física ($T = 1/\beta$).

Contudo, a expressão acima foi deduzida a partir do potencial fora do equilíbrio Φ_γ sem que impuséssemos quaisquer vínculos adicionais para garantir o equilíbrio termodinâmico. Uma forma de estabelecer as regiões de equilíbrio [56] é encontrarmos numericamente os valores de $\{\alpha, m, q\}$ que satisfaçam a equação de autoconsistência Eq. (2.25)

$$\left. \frac{\partial \Phi_\gamma}{\partial \alpha} \right|_\gamma = \varepsilon \Delta, \quad (5.13)$$

para valores de $\varepsilon \Delta$ e γ fixados. Depois, devemos selecionar no conjunto $\{\alpha, m, q\}$ destas soluções aquelas que minimizem $\Phi_\gamma(\varepsilon)$.

Pelo cálculo diferencial [5] sabemos que extremos de Φ_γ são detectáveis pelo teste das derivadas primeiras

$$\left. \frac{\partial \Phi_\gamma}{\partial m} \right|_{\varepsilon, \gamma} = 0, \quad (5.14)$$

$$\left. \frac{\partial \Phi_\gamma}{\partial q} \right|_{\varepsilon, \gamma} = 0. \quad (5.15)$$

Em seguida, devemos verificar quais destes extremos são mínimos de $\Phi_\gamma(\varepsilon)$, já que nesse caso a positividade da curvatura da hessiana do potencial generalizado é exigida

$$d_{\Phi_\gamma}(m, q) = \det \left(\begin{array}{cc} \frac{\partial^2 \Phi_\gamma}{\partial m^2} & \frac{\partial^2 \Phi_\gamma}{\partial m \partial q} \\ \frac{\partial^2 \Phi_\gamma}{\partial q \partial m} & \frac{\partial^2 \Phi_\gamma}{\partial q^2} \end{array} \right) \Big|_{\alpha, m, q} > 0. \quad (5.16)$$

Alternativamente, também deveríamos ser capazes de estabelecer critérios de equilíbrio termodinâmico diretamente sobre a entropia. Pois sabemos que no (limite $\gamma \rightarrow \infty$ do) ensemble microcanônico o equilíbrio termodinâmico implica a maximização da entropia s_{micro} . Enquanto no (limite $\gamma \rightarrow 0$ do) ensemble canônico é necessário à transformação de Legendre (TL), $F = U - TS$, da energia livre de Helmholtz (F) que S seja côncava [3].

Notemos que no EGE a entropia S_γ é definida pela Eq. (2.26) empregando-se uma transformação generalizada de Legendre do potencial Φ_γ , em analogia ao caso canônico. Portanto, se substituirmos a relação de autoconsistência [56] da Eq. (2.25) na entropia EGE em Eq. (2.26), poderemos definir uma *entropia EGE modificada* s_γ^* como

$$s_\gamma^* \doteq S_\gamma - \gamma \left(\frac{\partial \Phi_\gamma}{\partial \gamma} \right)_\alpha = \alpha U - \Phi_\gamma. \quad (5.17)$$

Dessa forma, s_γ^* é uma TL usual do potencial Φ_γ calculado no equilíbrio. Portanto, é lícito exigir a concavidade de s_γ^* como função de U . O que no caso da nossa solução pode-se escrever como

$$\frac{\partial^2 s_\gamma^*}{\partial \varepsilon^2} < 0. \quad (5.18)$$

Por outro lado, se fixarmos um intervalo de energias $\varepsilon \in (\varepsilon_{ini}, \varepsilon_{final})$ no qual exigimos a estabilidade termodinâmica, dada pela condição de equilíbrio Eq. (5.18), um valor mínimo $\tilde{\gamma}$ do parâmetro γ pode ser fixado univocamente. No caso específico do modelo BC, combinando as equações Eq. (2.26), (5.11), (5.12), (5.17) e (5.18), concluímos que

$$\gamma > \tilde{\gamma} \equiv \frac{-1}{2\Delta T^2(\varepsilon)c_\gamma(\varepsilon)}. \quad (5.19)$$

Desta forma sempre que $\gamma \geq \tilde{\gamma}$ podemos garantir a equivalência termodinâmica dos resultados microcanônicos e do EGE no intervalo $\varepsilon \in (\varepsilon_{ini}, \varepsilon_{final})$ [17, 18, 19, 20, 21, 22, 57]. O que permite-nos calcular o calor específico $c_\gamma(\varepsilon)$ segundo a definição

$$c_\gamma(\varepsilon) = \frac{d\varepsilon}{dT(\varepsilon)} = - \left. \frac{S_{\gamma,mm}}{T^2 d_{S_\gamma}(\varepsilon, m)} \right|_m \quad (5.20)$$

em analogia ao caso microcanônico [2], e também a susceptibilidade magnética

$$\chi_\gamma(\varepsilon, m) = - \frac{S_{\gamma,\varepsilon\varepsilon}/\Delta^2}{d_{S_\gamma}(\varepsilon, m)}. \quad (5.21)$$

Note que o determinante $d_{S_\gamma}(\varepsilon, m)$ da hessiana da entropia generalizada S_γ é escrito como

$$d_{S_\gamma}(\varepsilon, m) = \frac{1}{\Delta^2} \det \begin{pmatrix} S_{\gamma\varepsilon\varepsilon} & S_{\gamma\varepsilon m} \\ S_{\gamma m\varepsilon} & S_{\gamma mm} \end{pmatrix}. \quad (5.22)$$

Assim, no limite $\gamma \rightarrow \infty$ recuperamos a termodinâmica microcanônica do modelo, vide Figura (5.1), que para certos valores de (J, Δ, ε) tem diversas peculiaridades como:

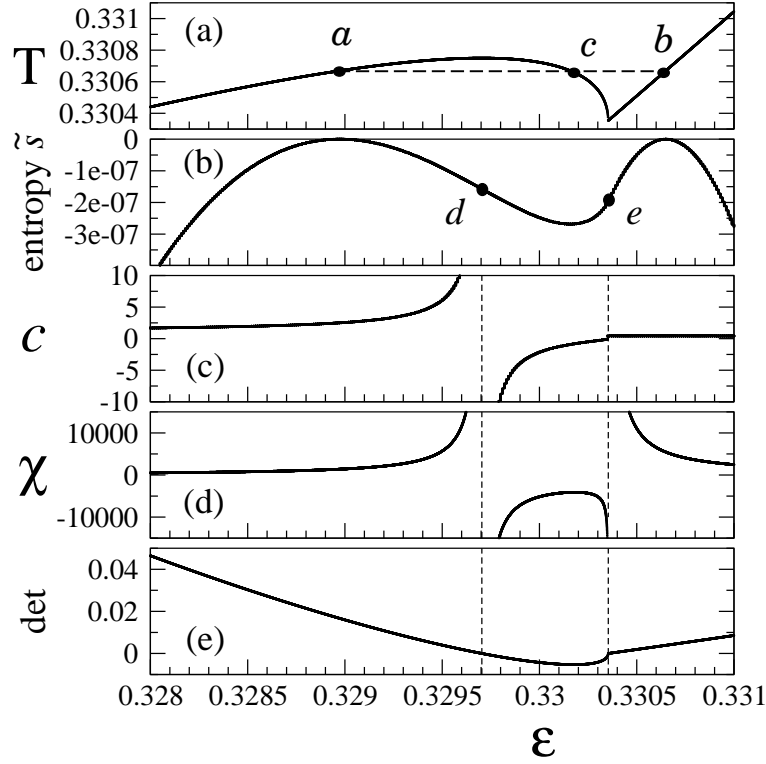


Figura 5.1: Comportamento no limite microcanônico da solução do modelo BC via formalismo EGE, com acoplamento $\Delta/J = 0.462407$, correspondendo à região de transição de fase de primeira ordem canônica e de segunda ordem microcanônica. (a) Temperatura microcanônica como função da energia média ε . A linha tracejada horizontal corresponde à temperatura crítica canônica. (b) A entropia deslocada $\tilde{s}(\varepsilon) = s_{micro}(\varepsilon) - (A + B\varepsilon)$, com $A = 0.401447$ e $B = 1.398397$. A subtração é efetuada para uma melhor visualização da não concavidade da entropia em relação a função linear ligando $s(\varepsilon_a)$ a $s(\varepsilon_b)$, com $\varepsilon_a = 0.328959$ e $\varepsilon_b = 0.330646$. (c) Calor específico $c(\varepsilon)$. Ele apresenta dois polos localizados pelos zeros do determinante $d_{S_{micro}}(\varepsilon, m)$, onde m denota os valores da magnetização que maximizam a entropia em dado ε . Esses pólos também podem ser observados a partir do comportamento de $T(\varepsilon)$ em (a). (d) Susceptibilidade magnética $\chi(\varepsilon)$. Ela apresenta dois polos, novamente posicionados nos zeros do determinante $d_{S_{micro}}(\varepsilon, m)$ e torna-se negativa entre eles. (e) Comportamento do determinante $d_{S_{micro}}(\varepsilon, m)$ como função de ε . As linhas tracejadas verticais mostram os zeros de $d_{S_{micro}}(\varepsilon, m)$.

não concavidades da entropia, calor específico e susceptibilidade magnética negativos e saltos, ou loops de van der Waals, na temperatura. Constatamos ainda haver regiões no diagrama de fases microcanônico em que ocorrem quebras de ergodicidade, as quais não estão associadas a transições de fase físicas, conforme apresentado na Figura (5.2).

Para valores intermediários da parametrização $0 < \gamma < \infty$, que fisicamente descrevem acoplamentos com banhos térmicos finitos, percebemos, Figura (5.3), como as curvas

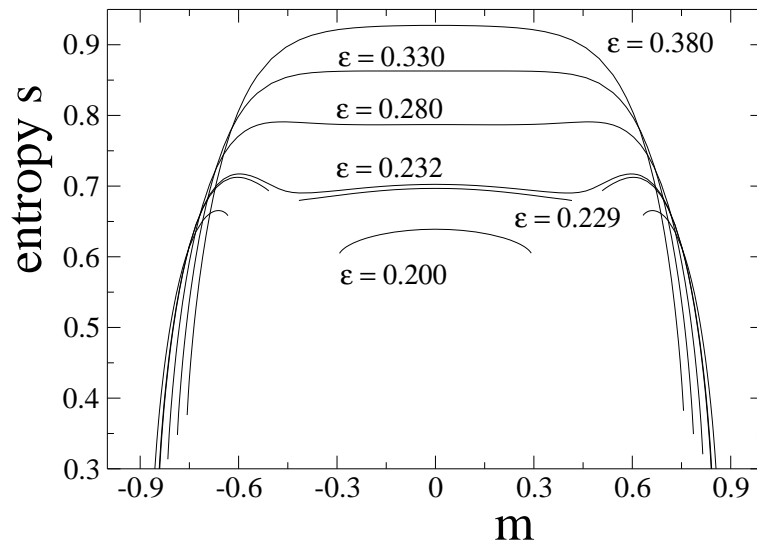


Figura 5.2: Entropia $s(\varepsilon, m)$ para alguns valores de ε e m com $\Delta/J = 0.462407$. Para valores inferiores da energia ε , alguns intervalos da magnetização são inacessíveis; o que demonstra quebra de ergodicidade neste modelo. Domínios magnéticos desconexos parecem ser típicos de sistemas com interações de longo alcance, exibindo transições de fase (canônica) de primeira ordem.

calóricas $T(\varepsilon) \times \varepsilon$ comportam-se com paulatinas variações em γ . Utilizando a relação Eq. (5.19) calculamos para (J, Δ) fixos o menor valor de γ que garante a equivalência termodinâmica² dos ensembles: canônico ($\tilde{\gamma} = 0$), EGE ($0 \leq \tilde{\gamma} < \infty$) e microcanônico ($\tilde{\gamma} \rightarrow \infty$) no intervalo $\varepsilon \in (\varepsilon_{ini}, \varepsilon_{final})$, vide Figura (5.4). É interessante notar ainda que regiões onde $\tilde{\gamma} \leq 0$, discutidas em [57], são interpretadas como fortemente acopladas (ou "super-acopladas") ao banho térmico.

De forma geral a análise da Eq. (5.22) constitui um poderoso discriminante de regiões onde pode haver equivalência de ensembles³. Por exemplo, observamos que funções resposta microcanônicas negativas aparecem em regiões onde $d_{S_{\gamma \rightarrow \infty}} \leq 0$, ou seja, são regiões microcanonicamente estáveis mas canonicamente metaestáveis ou instáveis. Sabe-se que nesses casos a entropia microcanônica deixa de ser localmente côncava e por isso as TL usuais são inaplicáveis. Portanto, o uso de ensembles generalizados, como o EGE, constitui uma alternativa à abordagem microcanônica para descrever metaestabilidades canônicas em situações de inequivalência, Figura (5.5).

²O que equivalentemente garante a concavidade local da entropia generalizada modificada s_{γ}^* .

³Sempre que há transição de fase canônica de primeira ordem pode ocorrer inequivalência de ensembles, desde que nesse caso a descrição microcanônica do fenômeno apresente funções resposta negativas. Por outro lado, quando há transições de fase de primeira ordem microcanônicas, i.e. com descontinuidades em $\beta(\varepsilon) \times \varepsilon$, a inequivalência de ensembles está garantida.

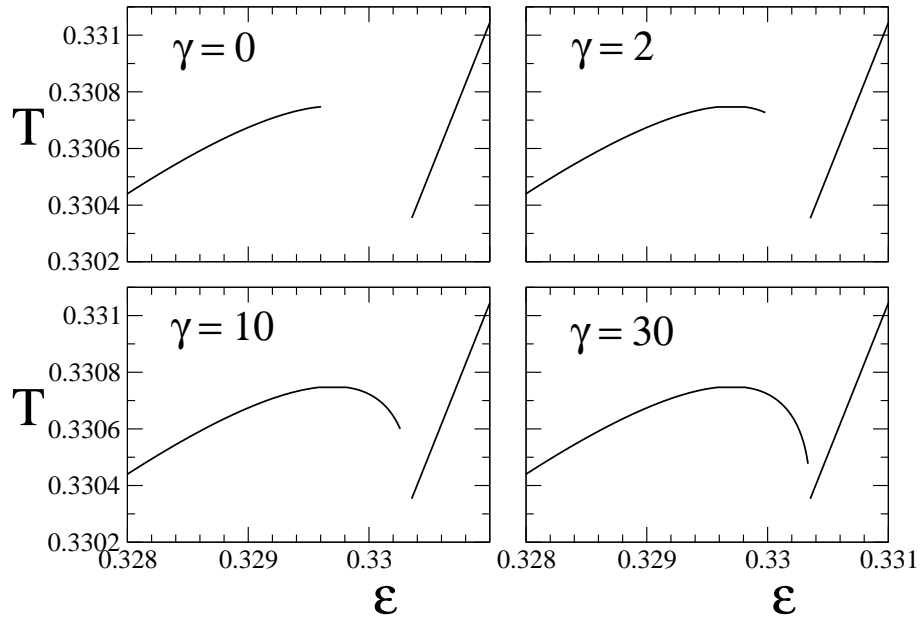


Figura 5.3: Curva calórica $T(\varepsilon) \times \varepsilon$ computadas para diversos valores de γ a partir da solução geral do modelo Blume-Capel no EGE. O acoplamento empregado é $\Delta/J = 0.462407$ e no limite de $\gamma \rightarrow \infty$ reobtemos a curva microcanônica (a) da Figura (5.1).

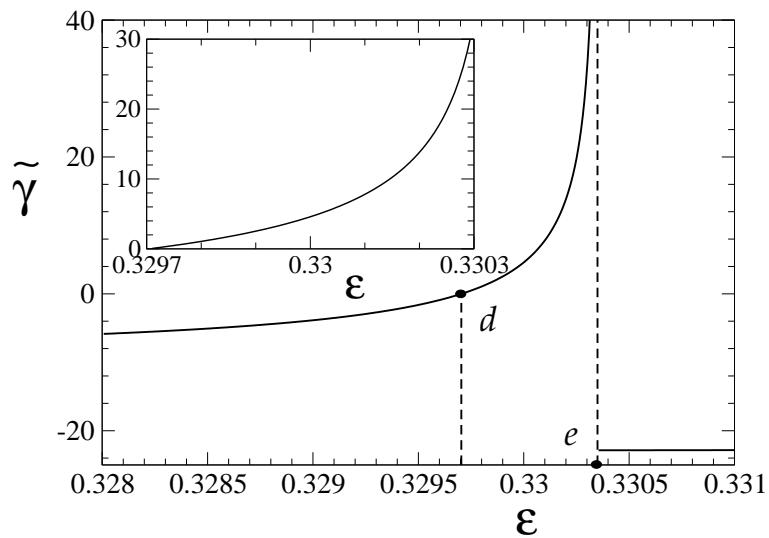


Figura 5.4: Valores mínimos que o parâmetro $\tilde{\gamma}$, calculado segundo a Eq. (5.19), deve assumir para que a termodinâmica deduzida via EGE para o modelo Blume-Capel (com $\Delta/J = 0.462407$) seja equivalente à microcanônica.

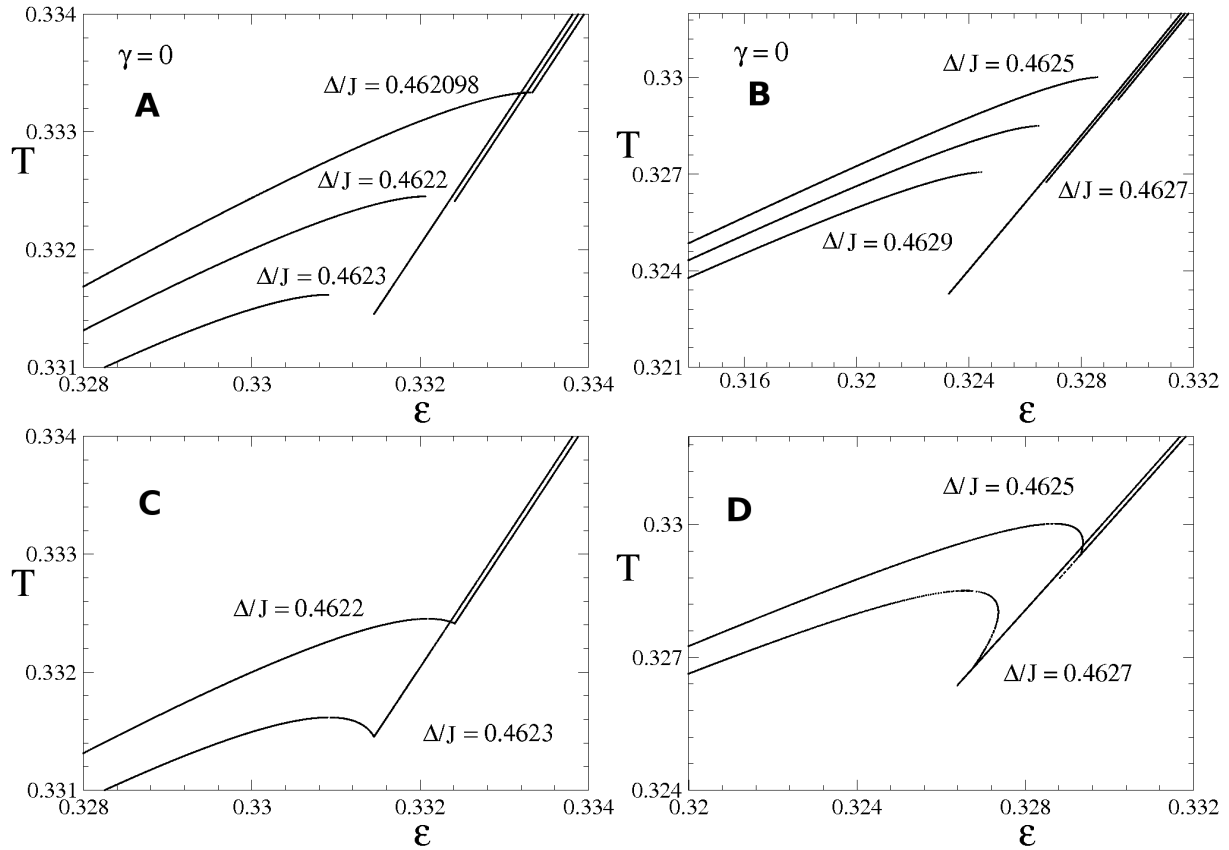


Figura 5.5: (A) Temperaturas EGE obtidas no limite canônico ($\gamma = 0$) para alguns valores de Δ/J . Para $\Delta/J = 0.462098$ temos simultaneamente transições de fase canônica e microcanônica de segunda ordem. Quando $\Delta/J = 0.4622$ e $\Delta/J = 0.4623$ as transições de fase canônicas são de primeira ordem, mas as microcanônicas são ainda de segunda ordem. (B) Temperaturas EGE no limite canônico ($\gamma = 0$), todos os valores de Δ/J estão na região de transição de primeira ordem canônica e microcanônica. (C) Temperaturas computadas via EGE no limite microcanônico ($\gamma \cong 10^{12}$), para $\Delta/J = 0.4622$ e $\Delta/J = 0.4623$ em que ocorrem transições de fase microcanônicas de segunda ordem e canônicas de primeira ordem. (D) Temperaturas EGE no limite microcanônico ($\gamma \cong 10^{12}$) na região de acoplamentos $\Delta/J = 0.4625$ e $\Delta/J = 0.4627$, cujas transição de fase são de primeira ordem.

5.4 Pontos tricríticos

Como vimos na seção anterior, o modelo BC apresenta em seu diagrama de fase $\Delta/J \times T_c$ dois pontos tricríticos distintos, um canônico e outro microcanônico. Estudamos aqui um sistema com forças de longo alcance, o qual apresenta inequivalência da termoestatística de ensembles, ou seja, seus diagramas de fase diferem, vide Figuras (5.6). Por exemplo, a transição de fase canônica de primeira ordem⁴ inicia-se na região em que $\Delta/J = 0.46209812$, enquanto microcanonicamente, a transição de primeira ordem⁵ localiza-se verdadeiramente em $\Delta/J = 0.46240788$.

A nossa presente abordagem produziu uma nova solução exata do modelo BC no ensemble gaussiano estendido, conforme mostra a Eq. (5.11). Podemos utilizar a relação Eq. (5.19) para *quantificar* a supracitada inequivalência, expressando-a como função de γ . Calculamos o valor mínimo que γ deve assumir, entre os dois pontos tricríticos, para que a termodinâmica do EGE seja equivalente à microcanônica vide na Figura (5.7).

Seria de esperar-se, segundo as expectativas teóricas mais conservadoras, que a solução canônica do modelo BC no ponto tricrítico em $\Delta/J = 0.46209812$ reproduziria a mesma termodinâmica computada microcanonicamente. Esta expectativa está correta e é reproduzida por nossa nova solução também para quaisquer valores assumidos pelo parâmetro γ . Contudo, como há inequivalência de ensembles neste modelo, existe um segundo ponto tricrítico distinto do primeiro, e que é encontrado apenas microcanonicamente quando o acoplamento é $\Delta/J = 0.46240788$. Nesta segunda situação, as soluções canônica e microcanônica são sabidamente inequivalentes, mas nossa solução interpolante é termodinamicamente equivalente à microcanônica se tomarmos $\gamma \geq 4950$. Por fim, o limite $\gamma \rightarrow \infty$ somente se faz obrigatório se quisermos estudar a região de transição de fase de *primeira ordem* microcanônica, i.e., $\Delta/J > 0.46240788$, utilizando o EGE.

Acreditamos que esta abordagem também possa ser útil na determinação numérica de pontos tricríticos onde haja inequivalência de ensembles. Para tal, pesos de Boltzmann-Gibbs poderiam ser substituídos por pesos gaussianos estendidos e variando-se γ em sucessivas simulações de Monte Carlo. Os resultados assim obtidos permitiriam construir diagramas de fase em diferentes ensembles para modelos sem soluções exatas.

⁴Caracterizada pela presença de calores latentes.

⁵Caracterizada pela presença de descontinuidades na temperatura

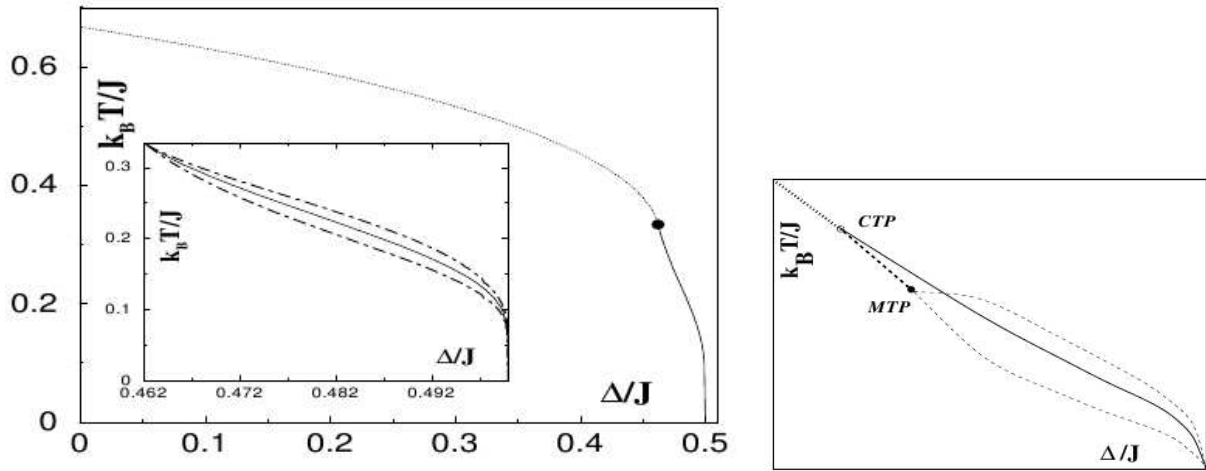


Figura 5.6: Na figura a esquerda temos linhas de transição de fase canônicas. A linha crítica (cinza pontilhada) termina no ponto tricrítico canônico •, a partir do qual a transição torna-se de primeira ordem (linha cheia). A figura menor é uma ampliação que mostra a linha (cheia) da transição de fase de primeira ordem canônica e as linhas (ponto-tracejadas) da transição de fase de primeira ordem microcanônica. Na figura a direita temos uma representação esquemática do diagrama de fase do modelo Blume-Capel, ampliado ao redor dos Pontos Tricrítico Canônico (CTP) e Microcanônico (MTP). A linha de transição de fase de segunda ordem (comum a ambos ensembles) é pontilhada; a linha canônica de primeira ordem é cheia e as linhas microcanônicas tracejadas são de segunda ordem em negrito e de primeira ordem em cinza. Adaptado da referência [10].

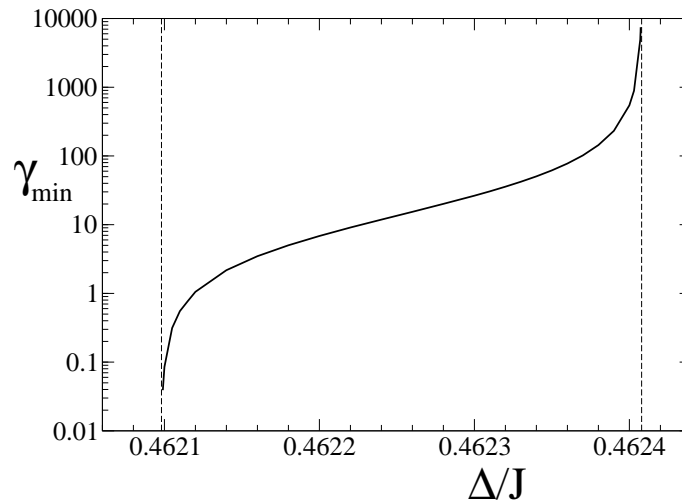


Figura 5.7: Valor mínimo do parâmetro interpolante γ para que a solução EGE seja termodinamicamente equivalente à microcanônica entre os pontos tricríticos canônico ($\Delta/J \cong 0.4621$) e microcanônico ($\Delta/J \cong 0.4624$).

Capítulo 6

Resultados numéricos: proteínas

Vos calculs sont corrects, mais votre physique est abominable
[A. EINSTEIN a G. Lemaître, 1927].

Apresentamos neste capítulo os detalhes técnicos e os resultados da implementação de nossas simulações microcanônicas empregando o algoritmo multicanônico e o modelo AB. Analisamos dois casos distintos: enovelamento de proteínas individuais e enovelamento/agregação¹ de duas proteínas interagentes. Foram estudados três tipos de sequências: Fibonacci (artificiais, com 13 aminoácidos), domínio Src SH3 (i.e., o peptídeo 1NLO com 56 resíduos) e Prion humana (i.e., a proteína 1HJM, com 104 resíduos). Empregamos neste estudo unidades reduzidas, com escala igual à distância de dois carbonos alfa consecutivos, conjuntamente com um modelo de célula esférica.

Inspirados pela metodologia desenvolvida em [45, 46] monitoramos as curvas calóricas e as funções resposta microcanônicas dessas biomoléculas visando caracterizar o seu comportamento termodinâmico. Esta abordagem permite quantificar a propensão biológica para o enovelamento [45] e agregação [43]. As barras de erro apresentadas em nossos cálculos provém da propagação de erro pelas fórmulas usuais. As curvas interpolantes mostradas em nossos gráficos foram obtidas usando filtros de média móvel visando a redução de ruídos (vide Apêndice D).

A análise dos resultados obtidos para todo o conjunto de proteínas é analisado à luz da física dos sistemas "pequenos" [2]. Traçamos ainda comparações entre a física proveniente destes sistemas biológicos e a do modelo BC estudado no capítulo anterior. Deste modo, os fenômenos bioquímicos de enovelamento e agregação protéicos são encarados como genuínas transições de fase microcanônicas.

¹Embora a literatura física não faça distinção a respeito, em bioquímica a agregação de poucas biomoléculas é mais conhecida como complexação protéica.

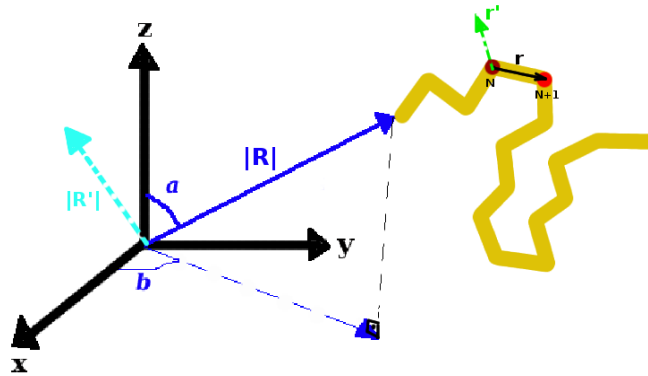


Figura 6.1: Esquema da atualização de Monte Carlo empregada em nossos estudos. Uma proteína posicionada a distância $|\vec{R}|$ da origem com ângulo polar "a" e equatorial "b" é atualizada para outra posição $\vec{R} \rightarrow \vec{R}' : \{|\vec{R}'|, a', b'\}$. Por outro lado, simultaneamente a posição relativa de cada um dos $(N + 1)$ -ésimos aminoácidos, em relação aos N -ésimos aminoácidos, é alterada de modo que $\vec{r} \rightarrow \vec{r}' = \vec{r} + \vec{dr}$.

6.1 Implementação computacional

Nos estudos que efetuamos do enovelamento e agregação protéicos, uma questão crucial foi como implementar eficientemente o processo de atualização das configurações moleculares. Isto porque, sistemas de proteínas apresentam barreiras nas suas energias livres dificilmente transponíveis mesmo utilizando o algoritmo MUCA. Daí a importância de empregarmos eficientes algoritmos de atualização configuracional que sofram o mínimo possível de aprisionamentos energéticos.

Para esse fim, aperfeiçoamos o conhecido método de atualização das calotas esféricas (*spherical-cap*, vide por exemplo [46]) introduzindo incrementos diminutos em ambos os ângulos diedrais. Em nossa versão, a qual permite simular diversas proteínas concomitantemente, efetuamos atualizações do tipo *pivot* [86, 87] para cada aminoácido escolhido aleatoriamente. A partir daí, a posição relativa do centro de massa \vec{R} de cada proteína é modificada por meio de uma dilatação (ou contração) seguida de rotação do seu vetor posição, vide Figura (6.1). Sendo a distância máxima permitida da proteína à origem igual $|\vec{R}| = 100$ unidades. Iniciamos as simulações com todas as proteínas distendidas e aleatoriamente posicionadas no espaço. Assim, o método de evolução que desenvolvemos é ergódico pois permite a visitação de todas as possíveis configurações no espaço de fase.

Para a estimativa dos parâmetros MUCA de cada uma das proteínas descritas a seguir empregamos $2 \times N \times 10^6$ atualizações angulares por repetição MUCA, considerando N como o número de aminoácidos em cada peptídeo. Cada uma das repetições MUCA com esse esquema de atualização demorou de 0,5h (1 Fibonacci) a 72h (2 Prions), perfazendo o tempo total simulacional de cerca de 250-dias-CPU em uma máquina Intel Core2 Quad de 3GHz.

6.2 Enovelamento e agregação

Esta seção é dedicada ao estudo do enovelamento e da agregação protéicas de três espécies distintas de peptídeos: uma sequência Fibonacci, artificialmente desenhada, com 13 aminoácidos [43, 44, 46], a sequência 1NLO do PDB correspondente ao domínio Src SH3 [48] e a Príon humana, de código 1HJM no PDB [41]. Quando analisadas individualmente, a única transição física averiguada é a de enovelamento; ao passo que analisando simulações envolvendo interação proteína-proteína constatamos também a presença da transição de agregação.

6.2.1 Sequências Fibonacci

As sequências do tipo Fibonacci foram idealizadas em [43, 44, 46]. Elas consistem de N monômeros com caráter hidrofóbico (A) ou hidrofílico (B), sendo geradas mediante o uso de determinadas relações de recorrência. Apesar de não estarem relacionadas a quaisquer proteínas naturais, as sequências Fibonacci são boas enoveladoras e agregadoras. Esta característica é bastante importante para estudos de natureza exploratória, pois nestes casos buscam-se exemplares cuja natureza permita entender qualitativamente os mecanismos envolvidos nas reais transições bioquímicas.

Além disso, validamos nossos códigos com um estudo exploratório de pequenas sequências desta família em dois casos complementares. No primeiro caso, empregamos cadeias individuais de tamanho $N = 13$ aminoácidos, os quais tiveram apenas seus ângulos diedrais atualizados, totalizando 26×10^6 passos de Monte Carlo por repetição MUCA. A respectiva análise microcanônica empregou os parâmetros multicanônicos obtidos nas 500 últimas repetições MUCA, em um universo de 1000 repetições efetuadas. A discretização da energia utilizada na construção dos histogramas empregou $\Delta E = 0,1$. O tempo total consumido nesta fase das simulações foi de cerca de 22h. A estrutura nativa, a de energia mínima encontrada, pode ser observada na Figura (6.2).

Na Figura (6.3) podemos ainda encontrar os resultados termodinâmicos computados a partir de um conjunto de parâmetros multicanônicos, com 50 estimativas equidistantes, determinadas no intervalo de repetições MUCA [500, 1000]. As barras de erro exibidas no gráfico da curva calórica, ou seja de $\beta(E) \times E$, correspondem ao desvio padrão deste conjunto de dados. À esta curva calórica sobrepusemos curvas interpolantes, produzidas pelo uso ($k - vezes$ recursivo) de filtros de médias móveis para m pontos. Ou seja, as curvas que na legenda da Figura (6.3) são denotadas por $k \times m$ ($m, k \in \mathbb{N}^*$) significam: $k - filtragens$ sucessivas efetuadas recursivamente com filtros de tamanho m . Tal abordagem foi necessária para o devido cômputo das derivadas $\frac{d\beta}{dE}$, que são exibidas na mesma Figura (6.3).

O cálculo de $C_V(E)$ é feito empregando-se a definição na Eq. (2.4), a qual escreve-se equivalentemente como $C_V = -\beta_{fit}^2 / \left(\frac{d\beta_{fit}}{dE} \right)$. Aqui utilizamos diretamente a curva $\beta_{fit}(E)$ que interpola os pontos em $\beta \times E$ e a sua derivada $\frac{d\beta_{fit}}{dE}$ calculada numericamente. Esta metodologia mostrou-se muito eficiente na redução de ruídos, além de indispensável na

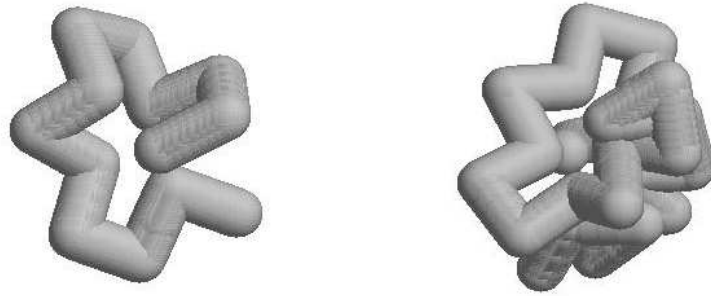


Figura 6.2: Estrutura nativa de *backbone* de sequências Fibonacci. (Painel esquerdo) configuração com energia mínima, $E_{1 \times Fib.} = -5,75$ e $\Delta E = 0,1$ para a sequência artificial com 13 resíduos $Fibonacci_{13}$: "ABBABBABABBAB". (Painel direito) configuração de energia mínima para a estrutura agregada $E_{2 \times Fib.} = -29,15$ e $\Delta E = 0,1$ de duas sequências idênticas de $Fibonacci_{13/AB}$.

estabilização do cálculo de $C_V(E)$. Nos resultados apresentados na Figura (6.3) visualiza-se claramente que há uma transição de fase contínua que interpretamos como enovelamento, embora não seja biologicamente rigoroso assim denominá-la. Esta transição é caracterizada por um único pico positivo no calor específico. A menor energia encontrada para o estado fundamental é $E_{ground} = -5,75$.

No segundo caso analisamos um conjunto de duas sequências Fibonacci com $N = 13$ e aplicamos o potencial de interação interprotéica da Eq. (4.3). Executamos também nessa etapa 1000 repetições MUCA. Cada repetição consistiu em 26×10^6 atualizações dos ângulos diedrais para cada sequência e 2×10^6 atualizações das suas posições de centro de massa. O tempo total desta segunda parte das simulações foi de cerca de 168h.

A análise dos parâmetros MUCA obtidos revelou a termodinâmica microcanônica deste sistema de forma idêntica ao caso de apenas uma Fibonacci. Empregamos aqui também 50 estimativas equidistantes dos parâmetros MUCA, coletadas dentre as 500 últimas repetições. Neste sistema observamos claramente uma transição de fase de primeira ordem (canônica), associada à agregação, cuja impressão digital é a presença de calores específicos negativos microcanônicos², Figura (6.4).

Além desta primeira transição, observada em energias maiores, notamos também indícios de picos positivos secundários em C_V . A presença de picos menores poderia estar associada a formação de domínios organizados do agregado, em estrita analogia aos verdadeiros processos biológicos aqui aproximados. Os resultados que obtivemos são compatíveis com os reportados nas referências [43, 44, 46].

²Muito embora essa transição seja canonicamente de primeira ordem, ela é microcanonicamente contínua, i.e. de segunda ordem, haja vista que não há saltos em $\beta \times E$

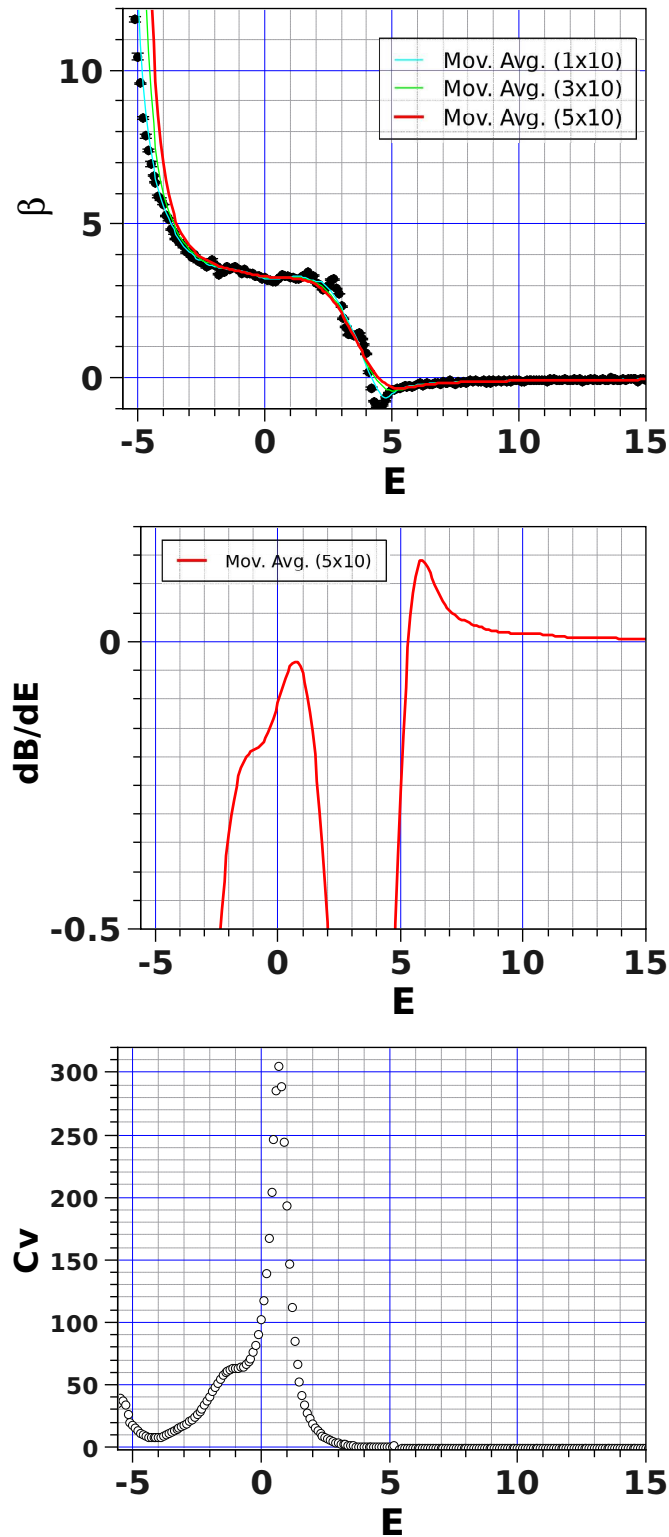


Figura 6.3: Termodinâmica microcanônica para uma única sequência Fibonacci. (Painel superior): curva calórica $\beta \times E$, barras de erro correspondem ao desvio padrão de 50 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas pela aplicação recursiva de k filtragens do tipo média móvel de 10 pontos. (Painel central): cálculo da derivada da curva interpolante (vermelha) de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se que ocorre um pico positivo caracterizando uma transição de fase contínua, ou de segunda ordem, associada ao "enovelamento".

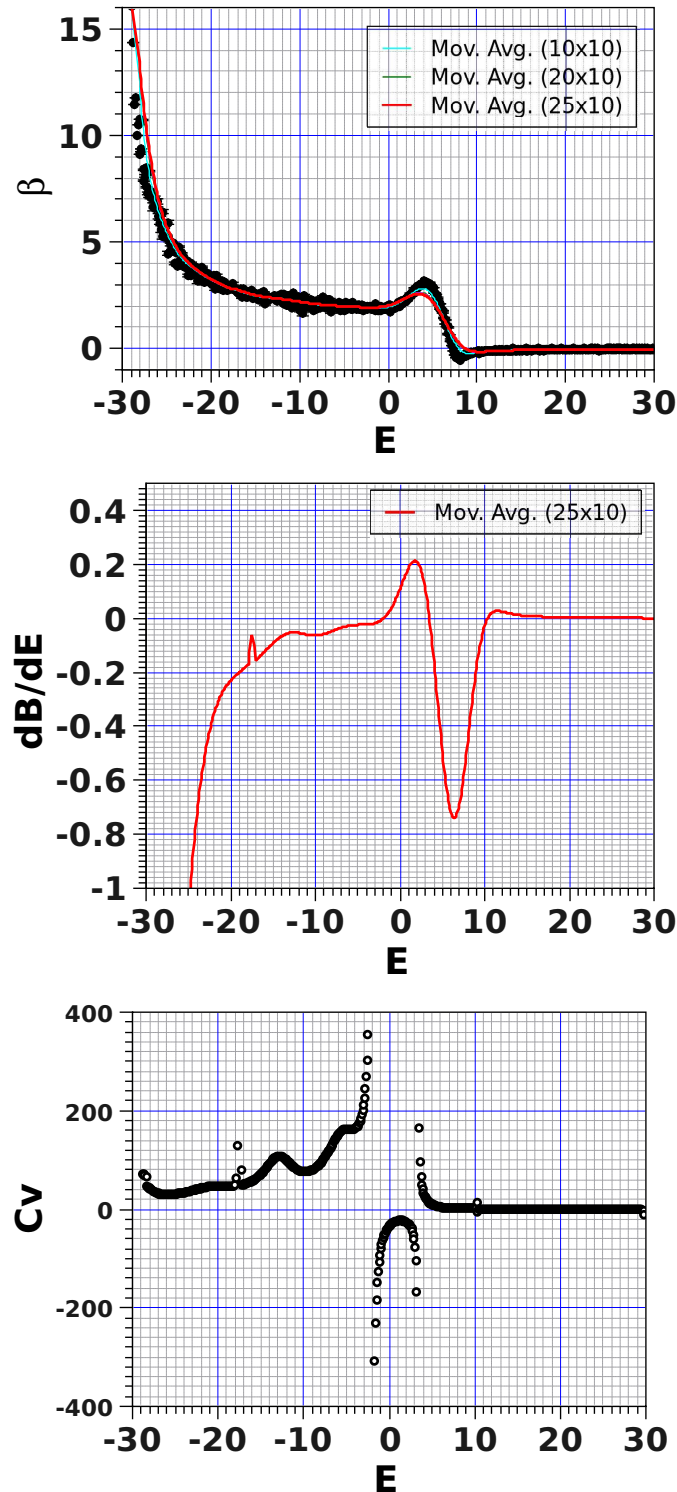


Figura 6.4: Termodinâmica microcanônica para duas sequências Fibonacci interagentes. (Painel superior): curva calórica $\beta(E) \times E$, barras de erro correspondem ao desvio padrão de 50 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas pela aplicação recursiva de k filtragens do tipo média móvel de 10 pontos. (Painel central): cálculo da derivada da curva interpolante (vermelha) de $\beta(E) \times E$, ou seja $\frac{d\beta(E)}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se uma transição de fase de primeira ordem (de agregação), com calores específicos negativos. Os pequenos picos positivos assinalam "formação de domínios" ou "enovelamento".

6.2.2 Domínios Src SH3

Na primeira aplicação biológica dos nossos códigos estudamos o domínio Src SH3 [48]. O domínio 3 de homologia Src, ou domínio SH3, é uma pequena proteína com 56 aminoácidos (código PDB: 1NLO) primariamente identificada como uma sequência conservada na proteína adaptadora viral v-Crk. Esta encontra-se também em partes não catalíticas de enzimas como as fosfolipases e diversas tirosinas citoplásmicas como as Abl e Src.

O domínio Src SH3, vide Figura (4.8), tem uma estrutura do tipo barril beta [24] contendo de cinco a seis fitas beta, apresentando um empacotamento denso de duas folhas beta antiparalelas. A região de ligação pode conter pequenas hélices. Cerca de 300 domínios SH3 são codificados pelo genoma humano. Os domínios SH3 clássicos [23, 24] também são encontrados em proteínas que interagem com outras proteínas e mediam a montagem de complexos específicos. Em humanos, estes domínios são restritos em geral a proteínas intracelulares.

Uma das mais interessantes características teóricas relativas a esses domínios foi apontada por recentes estudos empregando dinâmica molecular [48]. Observou-se que os domínios Src SH3 apesar de bons enoveladores formam agregados moleculares não patogênicos. Estes trabalhos especulam que a dinâmica envolvida nessas agregações mimetiza com alto grau de fidelidade àquela presente na formação de agregados mais complexos, como os de natureza priônica, servindo de excelente laboratório para o seu estudo.

Em nossas simulações mapeamos a estrutura primária do domínio Src SH3 em uma sequência do tipo AB segundo o caráter hidrofóbico dos seus aminoácidos (vide Capítulo 4). A sequência resultante foi "BAABABBBBAABBBBAAABBABAABAABBBBABAAABBAABBAB-BABABABBABBA". Em nossas simulações utilizamos duas diferentes discretizações das energias $\Delta E = 0,5$ e $\Delta E = 1,0$ para efeitos de comparação dos resultados MUCA. Encontramos pequena dependência quanto a essa discretização que parece influir apenas levemente na qualidade sinal/ruído dos parâmetros MUCA obtidos. Mostramos na Figura (6.5) as estruturas nativas para um único domínio Src SH3 e para seu estado dimerizado. Destacamos aqui a semelhança com as folhas beta vistas na Figura (4.8). Ilustramos os resultados termodinâmicos apenas para $\Delta E = 1,0$.

Para o cômputo dos parâmetros multicanônicos de uma sequência Src SH3 empregamos 3540 repetições MUCA. Cada repetição consistiu de $2 \times 56 \times 10^6$ atualizações dos ângulos diedrais. Dessas repetições utilizamos 54 conjuntos de parâmetros MUCA, distribuídos equidistantemente no intervalo das 540 últimas repetições, para a determinação da termodinâmica microcanônica deste peptídeo. Dessa forma, analisamos certamente apenas os parâmetros produzidos após o estado fundamental (com energia $E_{1 \times SH3} = -41,48$) já ter sido alcançado. Esta etapa das simulações demandou 142 dias-CPU.

Visando implementar a análise termodinâmica, ainda no caso de um único domínio Src SH3, foi necessário interpolar (recursivamente) entre os pontos da curva calórica calculada a partir de S , i.e. $[\beta \equiv \frac{dS}{dE}] \times E$: vide Figura (6.6). Utilizamos para tal diversas aplicações recursivas de um filtro de médias móveis de 5 pontos. Este tipo de análise estabiliza o cálculo das derivadas numéricas da entropia S e permite eliminar boa parte do ruído, que dificulta a caracterização das transições de fase. No caso de uma única proteína,

observamos claramente a presença de dois picos positivos no calor específico microcanônico, típicos de transições de segunda ordem. Estes picos podem ser relacionados biologicamente à formação de estruturas (e.g. das fitas beta) e em seguida ao seu enovelamento (pico menor), ocasionando a criação da estrutura nativa propriamente dita.

Na simulação da formação de agregados com dois domínios Src SH3 empregamos o potencial AB completo, dado na Eq. (4.3). Foram efetuadas 560 repetições MUCA, cada uma consistindo de $4 \times 56 \times 10^6$ atualizações dos ângulos diedrais e 2×10^6 atualizações das posições dos centros de massa. O tempo de processamento, para cada uma das discretizações energéticas foi de 187 dias-CPU. A energia mínima do estado fundamental mostrou-se fracamente dependente da discretização energética. Encontramos como estado fundamental $E_{2 \times SH3} = -29,15$ ($\Delta E = 1,0$).

As análises termodinâmicas efetuadas empregaram 26 configurações equidistantes, mensuradas a cada 10 no conjunto das repetições: [300,560]. Para a determinação das barras de erro usamos os desvios padrão simples. Já a estabilidade do cálculo das derivadas numéricas de $\beta \times E$ necessitou de diversas repetições recursivas de filtragens do tipo média móvel com 5 e 10 pontos. Estas curvas estão ilustradas na Figura (6.7) para mostrar o grau de dependência com o nível de filtragem empregado.

Ainda no caso da dimerização de Src SH3 observamos indícios de duas transições de fase fisicamente distintas. A primeira delas é claramente uma transição de fase (canônica) de primeira ordem³ com calor específico microcanônico negativo. Ela pode ser biologicamente interpretada como a transição de agregação (ou de dimerização) protéica e assinala a formação de estruturas fibrilares. Por sua vez, a localização e a caracterização da outra transição de fase são menos claras, dada a dependência com os tipos e graus de filtragem utilizados. Notamos certamente que ela é contínua e mostra pequenos picos no calor específico microcanônico. Estas propriedades podem ser indícios da formação de estruturas terciárias na Figura (6.5), isto é, mais complexas que um agregado protéico inicialmente amorfo como esquematizado na Figura (4.10). Notemos ainda que na Figura (6.7) poderíamos pensar que no intervalos de energias $E \in [-15, 30]$ estaríamos observando uma curva calórica típica, em plena concordância com as expectativas canônicas, mas a precisão dos nossos dados entre $E \in [20, 30]$ descarta essa possibilidade.

Os resultados aqui obtidos são bastante promissores, haja vista que a formação de agregados de domínios Src SH3 foi inicialmente prevista por estudos empregando dinâmica molecular [48] e potenciais interatômicos mais sofisticados que os nossos. Evidenciamos o poder do algoritmo MUCA para a análise de transições de fase microcanônicas em sistemas biológicos "pequenos". Além disso, o aparentemente simplificado modelo AB mostrou-se hábil em estender satisfatoriamente [88], ainda que em nível qualitativo, resultados teóricos prévios [48] destas complexas manifestações biológicas.

³Ao passo que como $\beta(E)$ é contínua, microcanonicamente esta transição também o é.

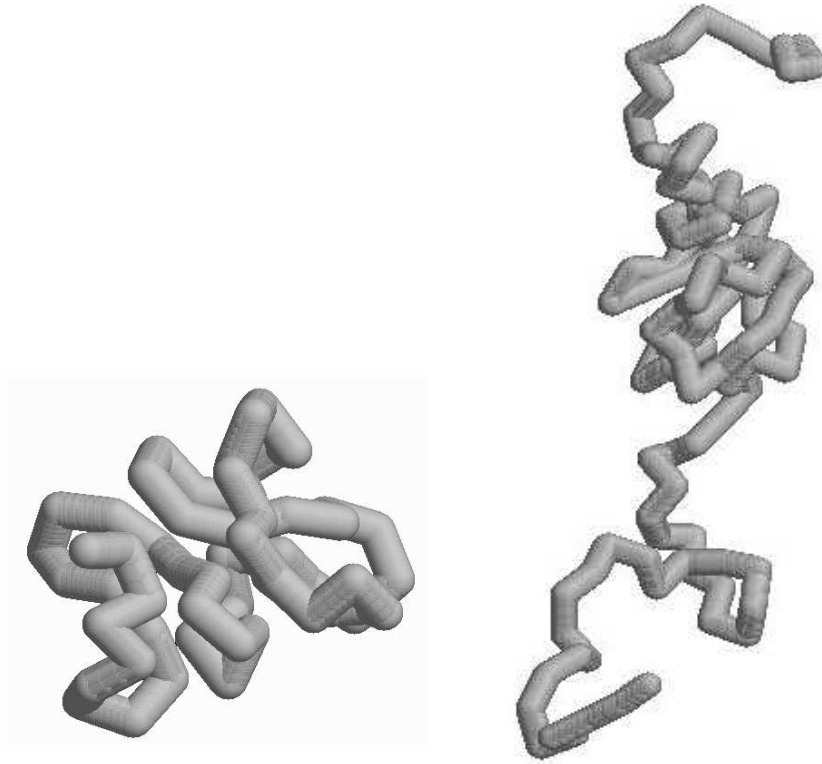


Figura 6.5: Estrutura nativa de *backbone* para domínios Src SH3 com 56 resíduos e código 1NLO no Protein Data Bank (PDB). Foi mapeada no modelo AB segundo sua natureza hidrofóbica/polar, resultando na sequência $SH3_{56/AB}$: "BAABABBBBAABBBBAAAABBABAABAABBBBABAABBAABBABBABABABBABBA". (Painel esquerdo): configuração nativa, ou com energia mínima $E_{1 \times SH3} = -41,48$ e $\Delta E = 1,0$ para a uma única sequência Src SH3. (Painel direito): configuração de energia mínima com $E_{2 \times SH3} = -29,15$ e $\Delta E = 1,0$ da forma agregada de dois peptídeos $SH3_{56/AB}$ interagentes.

6.2.3 Príons humanas

As proteínas 1HJM [40, 41] ou Príons humanas, podem ser mapeadas no modelo AB pela seguinte estrutura primária: "AAABAAAABAAAABAABAAABBBBABBABBABABBBBABBABABBBBABBABABBBBABBABABBBBABBABBBBABBABBBBABAABAAAABBAAABBBBABBAB". Estas proteínas por sua vez são notavelmente difíceis de simular. Não apenas por sua extensão molecular (104 resíduos) mas também por causa do seu potencial interatômico que produz um perfil altamente rugoso na energia livre. É esta estrutura energética peculiar que permite a essas proteínas formarem, de modo tão simples, agregados letais [32, 41, 73, 74].

Para tentar contornar essas dificuldades empregamos o já descrito algoritmo MUCA, com diversas discretizações energéticas $\Delta E = \{0.5, 1, \dots, 10\}$ em simulações exploratórias preliminares. O valor que permitiu mais rápida convergência para o estado fundamental e

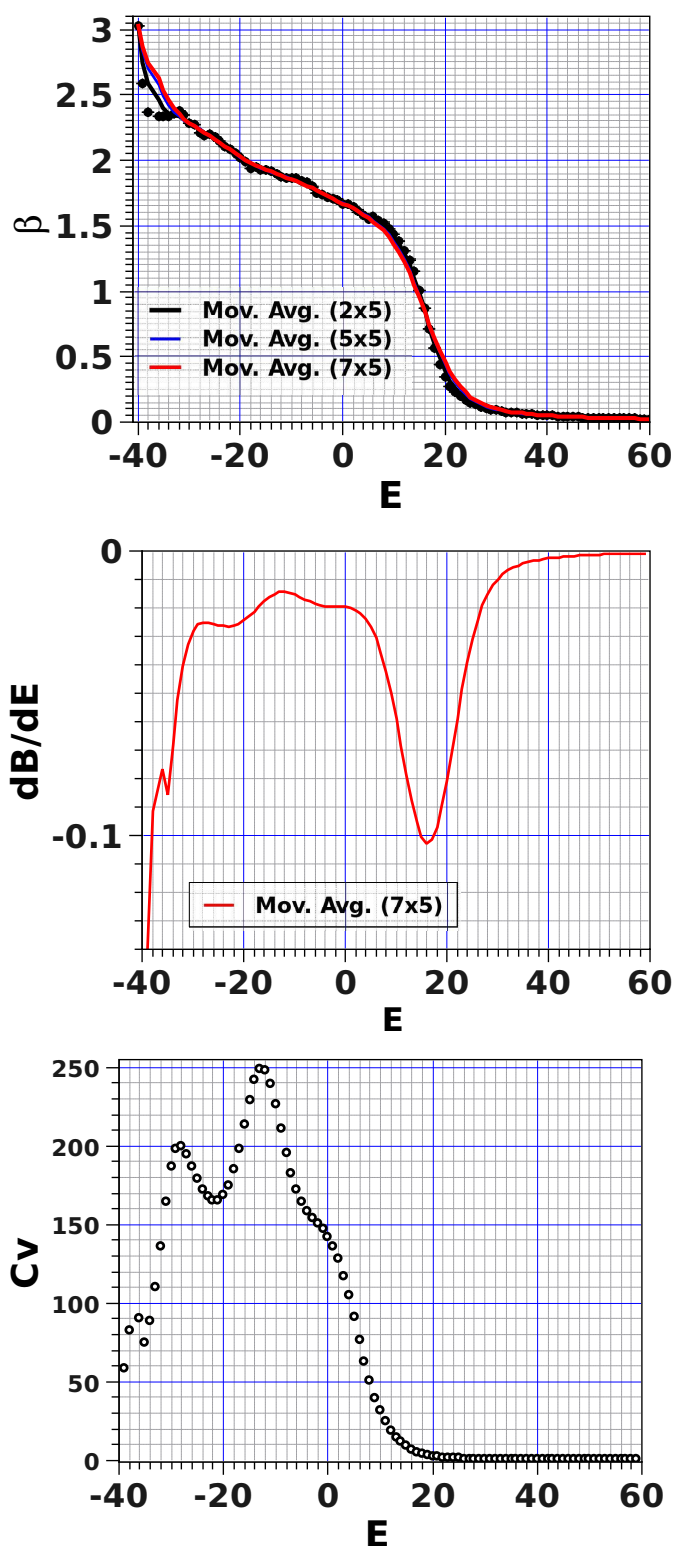


Figura 6.6: Termodinâmica microcanônica para um único domínio Src SH3. (Painel superior): curva calórica $\beta \times E$, barras de erro correspondem ao desvio padrão de 54 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas por aplicações recursivas de filtros de média móvel de 5 pontos. (Painel central): cálculo da derivada de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se que ocorrem dois picos positivos indicando transições de fase contínuas (ou de segunda ordem). Biologicamente assinalariam o envelhecimento (pico menor) e formação de estruturas secundárias (pico maior).

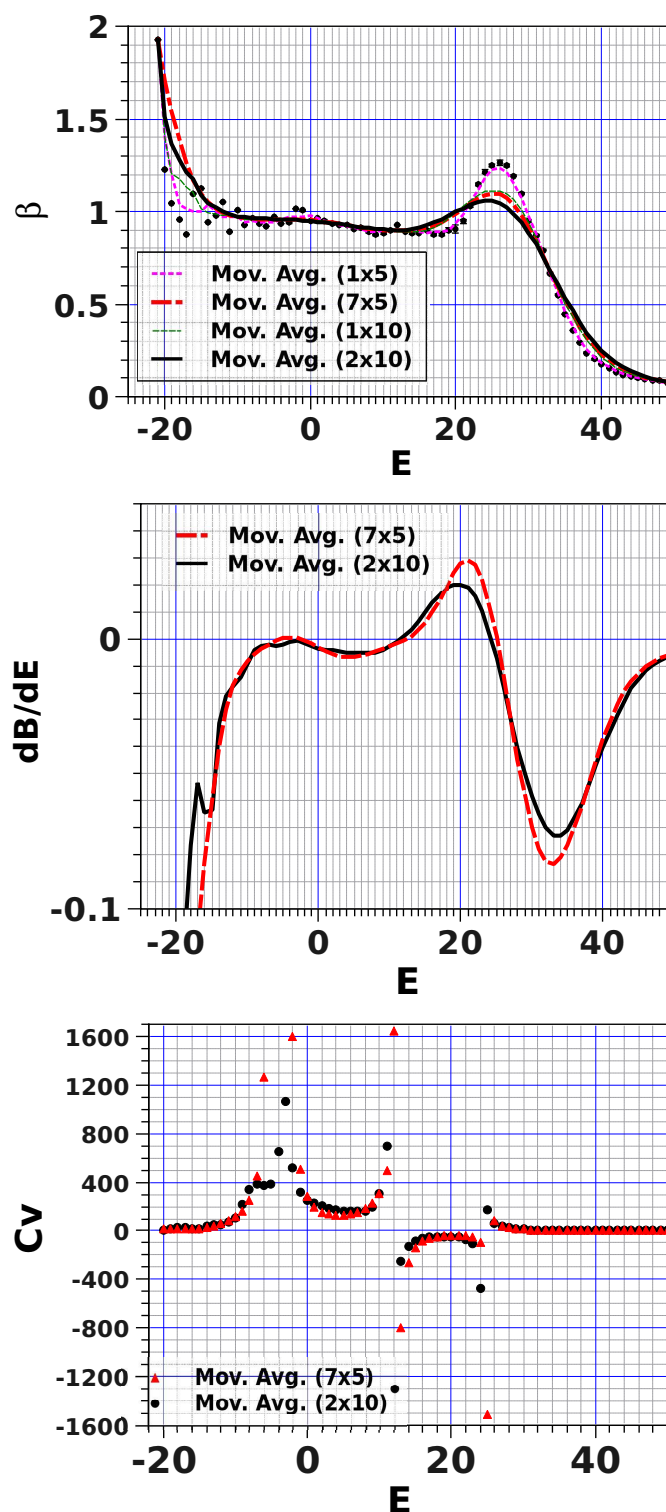


Figura 6.7: Termodinâmica microcanônica para dois domínios Src SH3 interagentes. (Painel Superior): curva calórica $\beta \times E$, barras vem do desvio padrão de 26 conjuntos de parâmetros MUCA. Linhas azul/vermelho emergem de 1 a 7 filtragens recursivas de médias móveis (de 5 pontos). (Painel Central): é a derivada da curva interpolante (vermelha) de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel Inferior): calor específico microcanônico do sistema ("filtro 7x5"). Há uma transição de primeira ordem (canônica, de agregação), com calores específicos microcanônicos negativos, e uma de segunda ordem (de envelamento).

mínimo ruído foi $\Delta E = 1, 0$. Após determinarmos este parâmetro computacional procedemos às simulações produtivas para uma e duas proteínas. O estado nativo da sequência AB que representa a proteína 1HJM tem energia mínima $E_{1 \times 1HJM} = -8, 35$. Esta sequência do tipo AB exibe uma forma enovelada, Figura (6.8), assemelhante à estrutura biológica real rica em hélices alfa como na Figura (4.11). Enquanto as formações fibrilares presentes em seus agregados têm energia mínima $E_{2 \times 1HJM} = 12, 59$ e podem ser vistos na Figura (6.8).

No caso de uma única Príon, pudemos executar 770 repetições MUCA com $104 \times 2 \times 10^6$ atualizações angulares em cada repetição. O tempo total desta simulação produtiva foi de 257 dias. Os parâmetros multicanônicos empregados para as devidas análises termodinâmicas provieram de 22 conjuntos, coletados a cada 10 repetições no intervalo [550,770]. A curva calórica $\beta \times E$ resultante não exibe saltos e tem como barras de erro os desvios padrão usuais, conforme mostrado na Figura (6.9). Foi preciso filtrar recursivamente $\beta \times E$ com o método de médias móveis de 5 pontos, isto garantiu bons níveis de sinal/ruído no cálculo de sua derivada. Assim, a partir das curvas interpolantes na Figura (6.9) obtivemos a derivada $\frac{d\beta}{dE} \times E$ com ela o calor específico $C_V \times E$.

A análise da curva calórica e do calor específico de uma única Príon evidenciam a presença de uma só transição de fase de segunda ordem. Este aspecto é importante pois biologicamente é associado a um "enovelamento sem formação de intermediários", como observado experimentalmente [73, 74]. Tal resultado comprova a eficácia do uso conjunto do algoritmo MUCA e do modelo AB pois capturaram, ainda que qualitativamente, características biológicas como a forma nativa e mecanismos de enovelamento também para a proteína 1HJM.

Por sua vez, no estudo da agregação de duas proteínas 1HJM empregamos em cada repetição MUCA $2 \times 208 \times 10^6$ atualizações angulares (diedrais) e 2×10^6 atualizações dos centros de massa protéicos. Nesta simulação efetuamos 80 repetições MUCA para as quais necessitamos de 240 dias-CPU. Analisamos 10 configurações multicanônicas no intervalo [70,80] e calculamos as barras de erro de $\beta \times E$ como mostra a Figura (6.10). Nesta curva calórica, aplicamos ainda até 10 filtragens recursivas de médias móveis de 5 pontos, o que permitiu o cômputo de $\frac{d\beta}{dE} \times E$ com uma boa relação sinal/ruído.

Apesar do massivo esforço para encontrarmos o estado fundamental deste sistema nos foi impossível obtê-lo com suficiente confiabilidade. Pois as menores energias encontradas são ainda positivas, não caracterizando portanto estados energeticamente ligados. Identificamos contudo, de forma clara e robusta, uma transição de fase de primeira ordem (canônica)⁴: a de agregação, vide $C_V \times E$ na Figura (6.10). Nenhuma outra transição de fase associada a formação de domínios ou interpretada como a de enovelamento foi observada, uma constatação potencialmente relevante à biologia destas moléculas, mas cuja confirmação futura dependerá de um maior poder computacional.

⁴Que a exemplo de todas as demais transições de agregação observadas neste estudo é microcanonicamente de segunda ordem, pois $\beta(E)$ é contínua também neste caso.

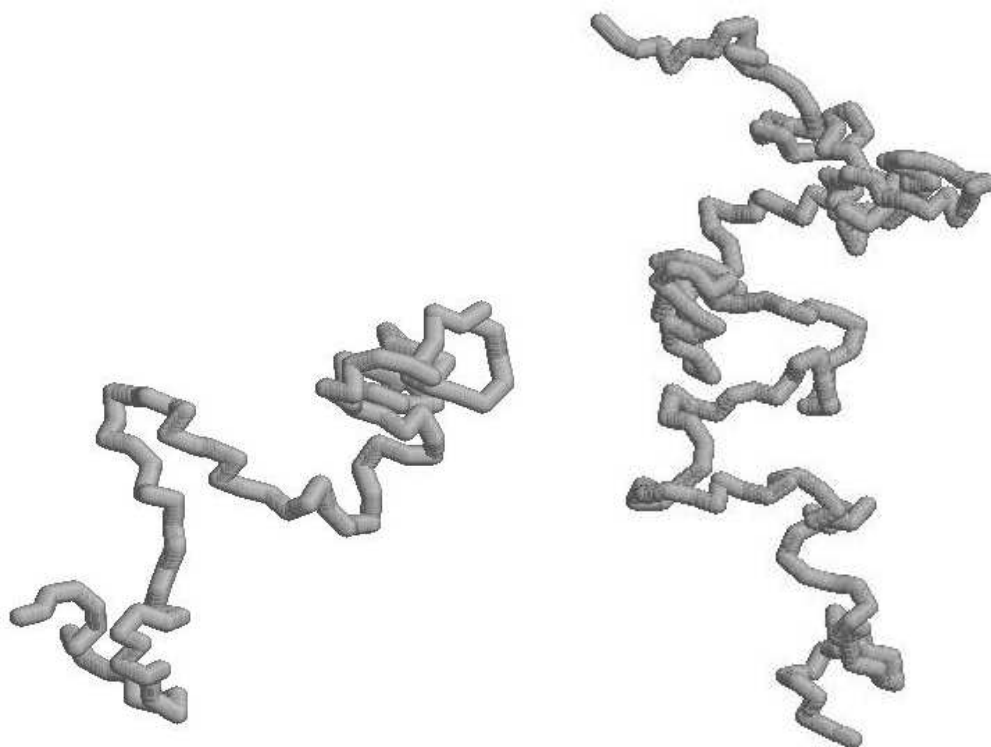


Figura 6.8: Estrutura de *backbone* de príons humanas com 104 resíduos e código 1HJM no Protein Data Bank (PDB). Foi mapeada no modelo AB segundo sua natureza hidrofóbica/polar, resultando na sequência $Prion_{104/AB}$: "AAABAAAABAAAABAABAAA BBBBABBABBABBBBBABBABBBBBABBBBAABBAABAB ABBBBABBBBBAB-BABBBBABAABAAABBAABBBBABBAB". (Painel esquerdo): configuração nativa, ou com energia mínima, $E = -8,35$ e $\Delta E = 1,0$ para a sequência $Prion_{104/AB}$. (Painel direito): configuração de energia mínima, dita na forma agregada ou dimerizada, com $E = 12,59$ e $\Delta E = 1,0$ para dois peptídeos $Prion_{104/AB}$ interagentes.

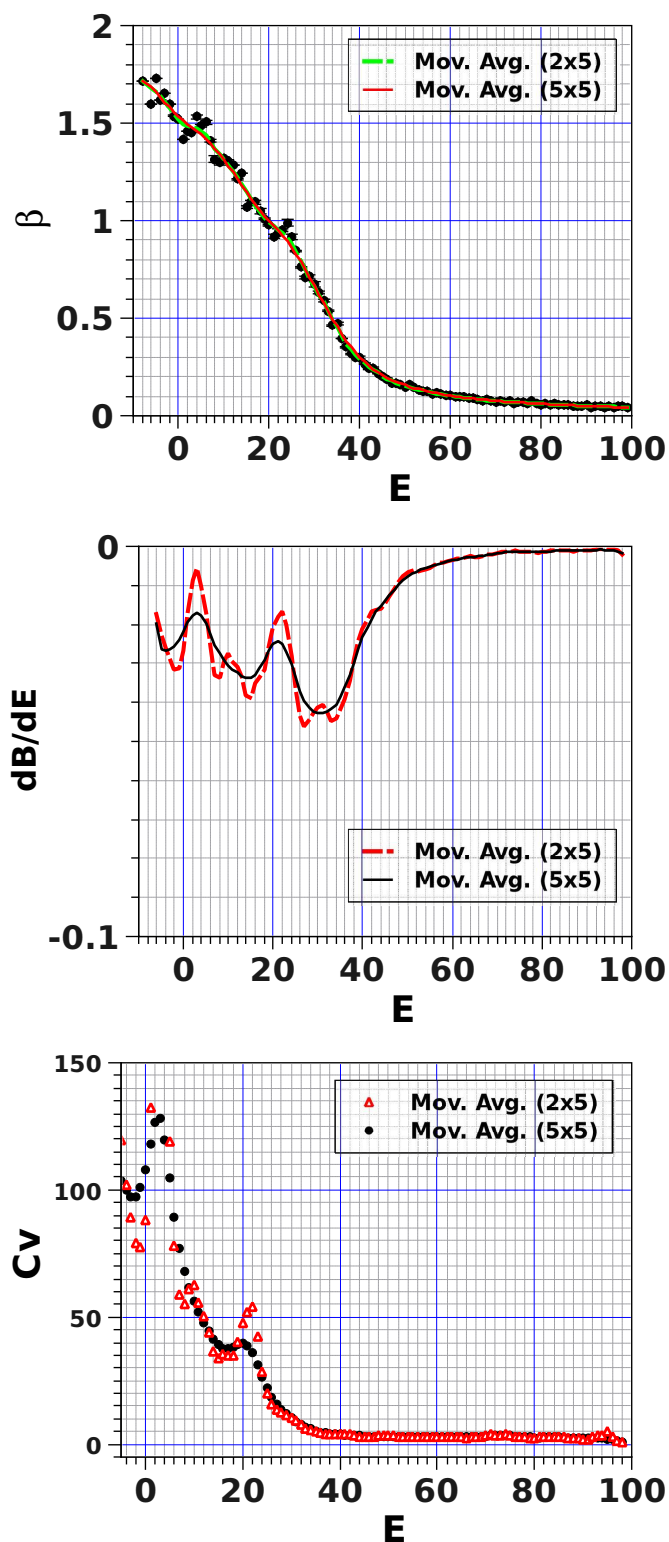


Figura 6.9: Termodinâmica microcanônica para uma proteína 1HJM, a Prion humana. (Painel superior): curva calórica $\beta \times E$, barras de erro correspondem ao desvio padrão de 22 conjuntos de parâmetros MUCA. As linhas contínuas são obtidas por filtragens recursivas do tipo média móvel de 5 pontos. (Painel central): cálculo da derivada de curvas interpolantes de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema, nota-se que ocorre um único pico positivo (estável), caracterizando uma transição de fase contínua (ou de segunda ordem), associada ao "enovelamento sem formação de intermediários": um fenômeno já observado *in vitro* [73, 74].

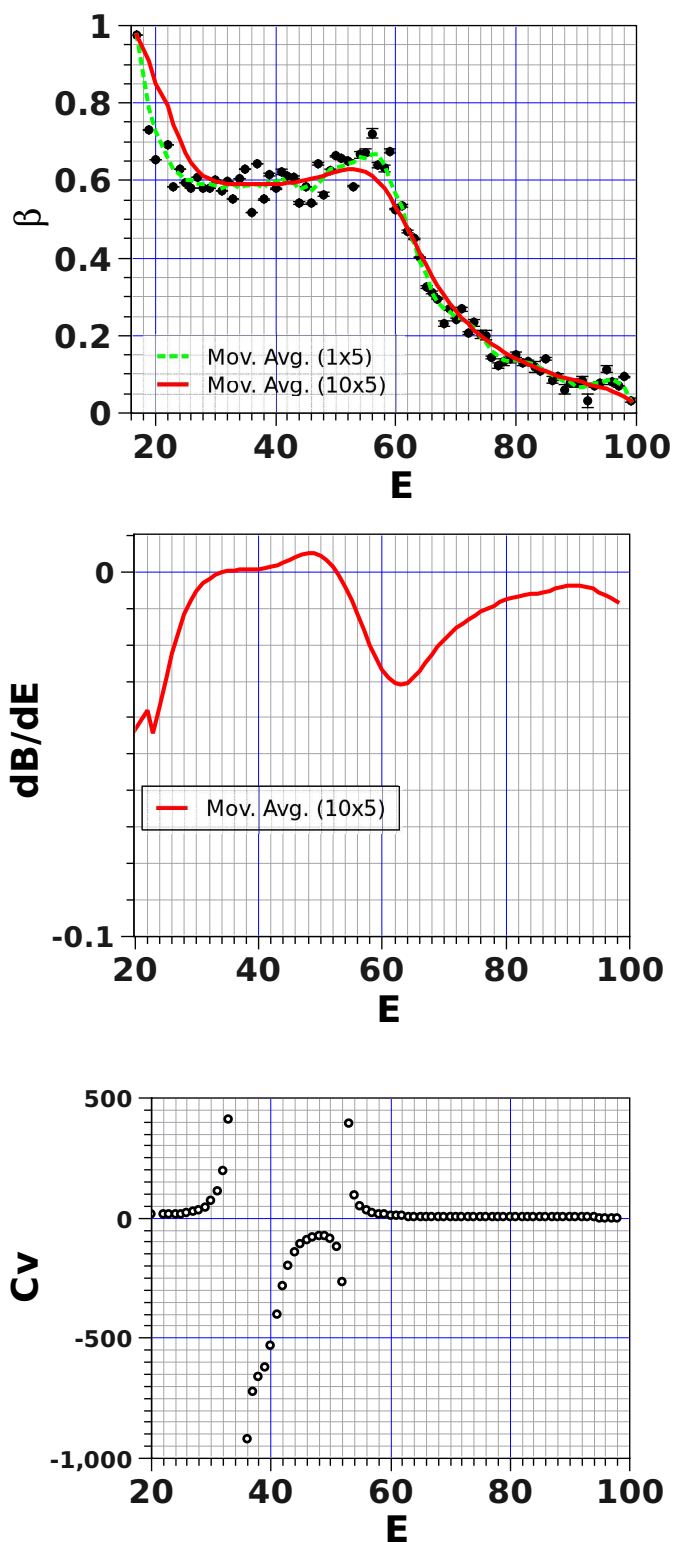


Figura 6.10: Termodinâmica microcanônica para duas proteínas 1HJM, ou Prions humanas. (Painel superior): curva calórica $\beta \times E$, barras de erro são o desvio padrão de 10 conjuntos de parâmetros MUCA. Linhas em verde/vermelho emergem de 1 a 10 filtragens recursivas de médias móveis (de 5 pontos). (Painel central): é a derivada da curva interpolante (vermelha) de $\beta \times E$, ou seja $\frac{d\beta}{dE} \times E$. (Painel inferior): calor específico microcanônico do sistema ("filtro 10x5"). Há apenas uma transição de primeira ordem (canônica, de agregação), com calores específicos negativos. Nenhuma outra transição de fase (e.g. contínua) foi observada, apesar de grandes esforços computacionais.

Capítulo 7

Considerações finais

Descubrí el secreto del mar meditando sobre una gota de rocío
[A. MACHADO].

Nesta tese apresentamos uma judiciosa revisão dos fundamentos da mecânica estatística microcanônica e tecemos considerações sobre como descrever e classificar transições de fase pela análise da entropia. Para tal, avaliam-se os autovalores e autovetores da matriz hessiana da entropia. Abordamos a formulação dos ensembles gaussiano estendido (EGE) e multicanônico (MUCA). Observamos ainda que para sistemas físicos "pequenos", cujos alcances característicos das interações equivalem aos tamanhos dos sistemas, as formulações da mecânica estatística são inequivalentes. Surge assim a questão de qual abordagem é a mais fundamental. Argumentamos em favor da microcanônica.

Neste contexto apresentamos nossa solução exata para o modelo Blume-Capel (BC) de alcance infinito [83] empregando o ensemble gaussiano estendido. Uma das principais características desse ensemble, o qual regulariza a formulação microcanônica, é que ele interpola suavemente os ensembles canônico e microcanônico. Portanto, uma solução no EGE é altamente desejável nos casos em que há inequivalência entre as soluções canônica e microcanônica, como no modelo BC. Por fim, introduzimos uma nova metodologia baseada no EGE em que o parâmetro interpolante γ é monitorado para localizarmos pontos tricríticos.

Mas, como a maioria dos sistemas fisicamente relevantes como as proteínas não admitem soluções exatas, se faz necessário o uso de métodos numéricos para estudá-los. Discutimos dentre estes métodos os algoritmos de Monte Carlo. O tópico mais sutil aqui mencionado é o fenômeno do frenamento crítico, no qual os tempos de autocorrelação integrados divergem nas vizinhanças de transições de fase, tornando as simulações ineficientes. Para contornar este problema o ensemble multicanônico oferece uma solução bastante atraente: diminui o frenamento crítico ao passo que implementa simulações microcanônicas.

Nossas investigações numéricas focalizaram o estudo mecânico-estatístico de proteínas. A modelagem teórica empregada neste trabalho baseou-se em potenciais interatômicos simplificados ditos de grão grosso. Por exemplo, no modelo AB adotado, apenas os potenciais

de torção entre aminoácidos e seu caráter hidrofóbico-hidrofílico foram considerados. Se por um lado ganhamos leveza computacional e habilidade em investigar a formação de agregados protéicos — e assim, os males degenerativos por eles causados — por outro, limitamo-nos a aspectos majoritariamente qualitativos destas transições de fase bioquímicas. Empreendemos então simulações de larga escala com recursos modestos¹, o que bastou para esclarecer importantes aspectos das transições de fase de enovelamento e agregação em três famílias de cadeias peptídicas.

A primeira dessas famílias engloba as chamadas cadeias Fibonacci, que não encontram paralelos na natureza. Até o momento, seu uso ficou restrito a estudos de cunho puramente teórico, cujos resultados permitiram depurar e otimizar nossos códigos. Para assegurar os resultados termodinâmicos foi necessário tratar cuidadosamente os parâmetros multicanônicos produzidos por nossas simulações. A saber: filtragens dos sinais resultantes visando melhorar a relação sinal/ruído mostraram-se indispensáveis para a adequada identificação das regiões de transição de fase. Ainda, o cômputo da curva calórica e das funções resposta termodinâmicas (C_V) evidenciaram o caráter de continuidade da transição de enovelamento. Além disso, esta abordagem mostrou que a agregação dessas cadeias é um fenômeno associado às transições de primeira ordem (canônicas), onde microcanonicamente temos $C_V < 0$. Esta é uma assinatura típica de sistemas "pequenos" em que há inequivalência de ensembles, exemplificados com o modelo BC.

Nossa primeira aplicação do modelo AB em sistemas biológicos veio do mapeamento do domínio Src SH3, cuja agregação já foi alvo de estudos prévios usando dinâmica molecular. Em paralelo, evidenciamos a grande robustez desta proposta quanto às escalas de hidrofobicidade adotadas². Assim, para a análise de cadeias individuais 3540 conjuntos de parâmetros multicanônicos foram produzidos, dos quais estudamos os últimos 540. Enquanto para entender sua dimerização (i.e. 2 proteínas em interação) computamos 560 conjuntos de parâmetros e analisamos o intervalo [300,560].

Observamos duas situações distintas com esta abordagem. Na primeira situação temos uma única proteína. A estrutura enovelada nativa, representada no modelo AB apenas pelos carbonos alfa (ou *backbone*), torna-se qualitativamente bastante semelhante àquela exibida no PDB. Neste caso, identificamos dois picos na curva do calor específico em função da energia, sinalizando transições de fase de segunda ordem. O primeiro pico no C_V pode estar biologicamente associado à formação de estruturas secundárias como folhas beta majoritariamente encontradas nesse peptídeo. O segundo pico está associado ao enovelamento. Esses resultados são numericamente estáveis e compatíveis com estudos prévios [48] que utilizaram outras metodologias, como a dinâmica molecular e potenciais Go.

Na segunda situação, em que foram simuladas duas cópias interagentes do domínio Src SH3, a física é distinta. Observamos claramente uma transição de primeira ordem (canônica) com calor específico microcanônico negativo, biologicamente associada a formação de

¹Totalizando cerca de 8 meses-CPU em uma máquina quadri-processada, Intel Core 2 Quad (Q9400), com códigos especialmente compilados para sua arquitetura (usando o Intel Fortran 11).

²Análise (preliminar) feita em conjunto com Matheus Mendonça, empregando o "Modelo de rede gaussiana ponderada", apresentada na VI CONFIAM 2010 (UNESP, Botucatu).

agregados. Uma segunda transição, contínua e que ocorre em energias menores, pode ainda ser localizada menos robustamente. Como $C_V > 0$ nesta transição, tendemos a associá-la à formação de domínios ou ao enovelamento. Lembremos que é esperado biologicamente que a formação dos agregados desordenados (ou protofibrilas) seja sucedida pelo aparecimento de estruturas mais organizadas, como as hélices beta-cruzadas, esquematizadas na Figura (4.10). Esta expectativa está de acordo com nossas simulações.

Estudamos também a formação de estruturas enoveladas e agregadas a partir da proteína humana 1HJM mapeada no modelo AB. Para tal, necessitamos de cerca de 8 meses de simulações ininterruptas. Assim, a necessidade de computação paralela em futuros estudos ficou evidente. Para uma proteína isolada, das 770 configurações de parâmetros MUCA obtidas, analisamos o intervalo [550, 770] que seguiu-se à determinação do estado fundamental. Curiosamente, se comparada às transições das sequências Fibonacci e Src SH3, observamos no caso das Prions apenas um pico no calor específico associado a uma transição contínua: a de enovelamento protéico. Esta observação está de acordo com evidências experimentais [73, 74] que sugerem que Prions enovelam-se sem exibir estados intermediários.

Simulamos também a dimerização de proteínas 1HJM, com a produção de 80 conjuntos de parâmetros MUCA, totalizando 240 dias-CPU. Não tivemos sinais críveis da obtenção de um estado fundamental com energia mínima já estabilizada. Mesmo assim, procedemos à análise termodinâmica com os últimos 10 conjuntos de parâmetros MUCA. Apesar da evidente necessidade de filtragens para melhorar a relação sinal/ruído, identificamos claramente a transição de fase de agregação como de primeira ordem (canônica). Ainda, diferentemente das demais proteínas aqui estudadas, notamos a ausência de quaisquer outras transições de fase. Entretanto, nossos dados não permitem excluir a existência de uma transição de fase (contínua) adicional, que poderia ser biologicamente relevante. Simulações futuras empregando maior poder computacional serão necessárias para esclarecer este tópico.

Por meio de nossas investigações numéricas concluímos que a formação de estruturas terciárias em proteínas está associada a transições de segunda ordem. Por outro lado as transições de agregação associam-se a transições de fase microcanonicamente contínuas (com calores específicos negativos), mas canonicamente de primeira ordem, isto é, com presença de calor latente. Um exemplo claro de inequivalência de ensembles em sistemas biomoleculares. A existência de mecanismos universais atuantes nos processos de enovelamento e agregação protéicos, como já sugerido por investigações prévias empregando dinâmica molecular [48], é uma hipótese coerente também com nossa análise. Portanto, o modelo minimalista AB que utilizamos nesta tese parece descrever corretamente características biologicamente importantes a custos computacionais aceitáveis.

Por fim, uma futura adoção de outros potenciais de grão grosso [71] e de algoritmos paralelos [89] poderá permitir descrições com maior nível de riqueza de detalhes estes sistemas protéicos. Uma vez implementados, as perspectivas para estes estudos são promissoras. Especialmente quando for possível analisar a formação de agregados na presença de enzimas antagonistas [91], uma aplicação *in silico* desejável ao desenvolvimento de novas terapias contra proteinopatias degenerativas.

Apêndices

It remains that, from the same principles, I now demonstrate the frame of the System of the World.

[SIR ISAAAC NEWTON, Principia].

Apresentamos estes apêndices para efeito de maior completude teória desta tese. Não visamos o rigor matemático, mas sim, fornecer uma simples revisão seguindo de perto a literatura [3, 5] sobre alguns tópicos e técnicas úteis à compreensão do presente estudo.

Apêndicê A

O método *steepest-descent*

Suponha que queiramos calcular uma integral de contorno da forma

$$I \equiv \int_{\gamma} e^g dz,$$

onde γ é um contorno no plano complexo e $g(z)$ é uma função analítica complexa de z . Um valor aproximado para I pode ser obtido como segue.

Admitimos que o módulo do integrando é máximo em um ponto z_s em que $u(z) \equiv \text{Re}(g(z))$ tem um pico. Nesse ponto $\partial u/\partial x = \partial u/\partial y = 0$, onde $z = x + iy$. Pelas condições de Cauchy-Riemann em z_s , $\text{Im}(g)$ também tem derivadas nulas em relação a x e y . Assim temos $dg/gz = 0$ em z_s . Consequentemente, a expansão de g em série de Taylor é dada por

$$g(z) = g(z_s) + \frac{1}{2!}g''(z_s)(z - z_s)^2 + \dots,$$

e podemos escrever $g''(z_s) = -2ae^{i\phi}$ e $z - z_s = \epsilon e^{-i\phi/2}$, onde $a > 0$ e ϵ, ϕ e ϕ' são números reais. Então

$$g(z) \approx g(z_s) - ae^{i(\phi-\phi')}\epsilon^2,$$

Vamos agora deformar o contorno original de integração γ até que ele passe por z_s ao longo da linha $\phi' = \phi$. A contribuição para a integral I dos pontos próximos de z_s podem agora ser escritos como

$$I' \equiv e^{-i\phi/2} e^{g(z_s)} \int_{\epsilon\text{-pequeno}} e^{-a\epsilon^2} d\epsilon. \quad (7.1)$$

Embora a expansão em Taylor em que essa expressão se baseia seja válida somente para valores de ϵ pequenos, a natureza Gaussiana do integrando da Eq. (7.1) nos permite estender os limites de integração para $\pm\infty$ sem erro apreciável. Assim, da integral gaussiana usual, podemos escrever

$$I' \simeq e^{-i\phi/2} e^{g(z_s)} \sqrt{\pi/a}.$$

O método do *steepest-descent* consiste em equacionar I com a contribuição de I' a partir das vizinhanças de z_s , isto é

$$\int_{\gamma} e^g dz \simeq e^{i\theta} e^{g(z_s)} \sqrt{\frac{2\pi}{|g''(z_s)|}},$$

onde $\theta \equiv -\frac{1}{2} \arg[-g''(z_s)]$. A versão no domínio real deste método é chamada de método do ponto de sela.

Apêndice B

A transformada de Laplace

A transformada de Laplace de uma função $f(E)$, definida para todos os números reais $E \geq 0$, é uma função $F(\beta)$ definida por

$$F(\beta) = \mathcal{L}\{f(E)\} = \int_0^{\infty} e^{-\beta E} f(E) dE,$$

onde o parâmetro β é um número complexo $\beta = \sigma + i\omega$, com σ e ω reais.

O significado da integral depende do tipo das funções de interesse. Uma condição necessária para a existência destas integrais é que f deve ser localmente integrável no intervalo $[0, \infty)$. Para funções localmente integráveis que decaem a zero apenas no infinito, a integral pode ser entendida como uma integral própria de Lebesgue.

A transformada inversa de Laplace

Podemos calcular a transformada inversa de Laplace por meio de uma integral complexa chamada de integral de Bromwich, Integral de Fourier-Mellin ou fórmula inversa de Mellin

$$f(E) = \mathcal{L}^{-1}\{F(\beta)\} = \frac{1}{2\pi i} \lim_{T \rightarrow \infty} \int_{\gamma-iT}^{\gamma+iT} e^{\beta E} F(\beta) d\beta,$$

em que γ é um número real tal que o caminho de integração está na região de convergência de $F(\beta)$, normalmente requerendo $\gamma > \text{Re}(\beta)$ para toda singularidade em $F(\beta)$. Se todas as singularidades estão no semiplano à esquerda, ou seja $\text{Re}(\beta) < 0$ para todo β , então γ pode ser tomado como zero e a fórmula da integral inversa acima torna-se idêntica à transformada inversa de Fourier.

Apêndice C

As transformações de Legendre

Em matemática, as transformações de Legendre (TL) são operações que transformam uma função real de uma variável em outra. A função $f(x)$ é transformada em $f^*(p)$ pela definição

$$f^*(p) = \max_x (px - f(x)).$$

Se $f(x)$ é diferenciável, então $f^*(p)$ pode ser interpretada como o negativo do intercepto em y da reta tangente ao gráfico de $f(x)$, cuja inclinação é p . Em especial, o valor de x que produz um máximo é $f'(x) = p$. Isto é, a derivada da função $f(x)$ torna-se o argumento da função $f^*(p)$. No caso em que $f(x)$ é convexa, $f^*(p)$ satisfaz $f^*[f'(x)] = xf'(x) - f(x)$.

A transformação de Legendre é sua própria inversa, como ocorre com as transformadas de Fourier. Mas, enquanto a transformada de Fourier consiste de uma integração com um núcleo, a transformação de Legendre usa *maximização*. Esta transformação é bem comportada apenas se $f(x)$ for convexa.

Em outras palavras, as TL são uma aplicação das relações de dualidade existentes entre pontos e retas. Assim, as relações funcionais especificadas podem ser representadas igualmente bem por um conjunto de pontos (x, y) , ou por um conjunto de retas tangentes, especificadas por suas inclinações e interceptos.

Ilustração termodinâmica

A estratégia das TL é transformar uma função em outra, que dependerá de uma nova variável correspondendo a derivada parcial da função original com relação à variável independente.

Por exemplo, enquanto a energia interna U é uma função explícita de variáveis extensivas como a entropia e o volume,

$$U = U(S, V, \{N_i\}),$$

a entalpia H trata-se de uma transformada de Legendre de U com relação a PV ,

$$H = U + PV = H(S, P, \{N_i\}), \quad P = -\left(\frac{\partial U}{\partial V}\right)_S,$$

a qual torna-se uma função da entropia e de uma quantidade intensiva (a pressão). Estas são as novas variáveis naturais úteis quando a pressão externa P é constante. As energias

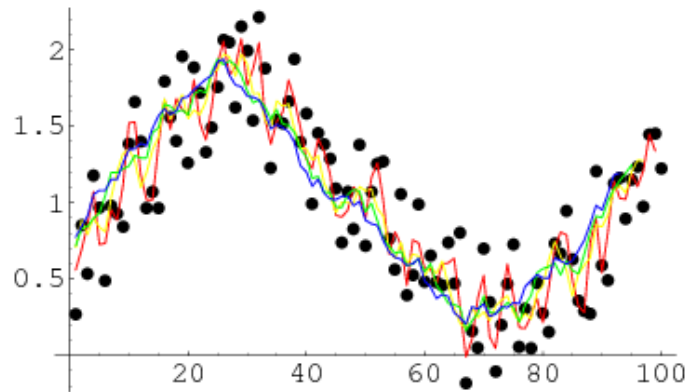


Figura 7.1: O gráfico acima mostra um conjunto de dados (100 pontos, em preto) e suas médias móveis com tamanhos diferentes [5]. Temos respectivamente as seguintes curvas: 2-pontos (vermelha), 4-pontos (amarela), 6-pontos (verde) e 8-pontos (azul). Note o efeito de filtragem do ruído nestas curvas.

livres (de Helmholtz e Gibbs) são também obtidas através de transformações de Legendre sucessivas, subtraindo-se TS a partir de U e de H , respectivamente.

Apêndice D

O filtro de média móvel

Dada uma sequência $\{a_i\}_{i=1}^N$, uma n -média móvel é uma sequência $\{s_i\}_{i=1}^{N-n+1}$ definida a partir de $\{a_i\}$ tomando-se a média das subsequências de n -termos [5], ou seja

$$s_i = \frac{1}{n} \sum_{j=i}^{i+n-1} a_j. \quad (7.2)$$

Assim, as sequências S_n dadas pelas n -médias móveis são dadas, por exemplo, por

$$S_2 = \frac{1}{2} (a_1 + a_2, a_2 + a_3, \dots, a_{n-1} + a_n),$$

$$S_3 = \frac{1}{3} (a_1 + a_2 + a_3, a_2 + a_3 + a_4, \dots, a_{n-2} + a_{n-1} + a_n).$$

Os efeitos deste tipo de filtro podem ser vistos na Figura (7.1) para diferentes quantidades de pontos empregados para tirar a média na Eq. (7.2).

Apêndice E

O ensemble canônico

Apresentamos nesta seção duas possíveis formas de deduzir o ensemble canônico a partir da maximização vinculada da entropia ou termodinamicamente a partir de um sistema acoplado a um reservatório térmico.

Dedução a partir da teoria de ensembles

Seja E_i a energia do microestado i , suponha que existam n_i partículas ocupando este estado. Assumimos também que o número de partículas no ensemble é \mathcal{N} e sua energia é \mathcal{E} , ambos são fixados de acordo com

$$\begin{aligned}\mathcal{N} &= \sum_i n_i, \\ \mathcal{E} &= \sum_i n_i E_i.\end{aligned}\tag{7.3}$$

Uma vez que as partículas são indistinguíveis, para cada conjunto $\{n_i\}$ o número de possíveis modos de embaralhar o sistema, que ainda assim especificam o mesmo estado macroscópico do ensemble, é igual a

$$W(\{n_i\}) = \mathcal{N}! / \prod_i n_i!\tag{7.4}$$

A distribuição mais provável é aquela que maximiza a Eq. (7.4), mas respeitando os vínculos da Eq. (7.3). Assim, a probabilidade para que as outras distribuições ocorram será muito pequena ao tomarmos o limite termodinâmico $\mathcal{N} \rightarrow \infty$. Para determinarmos a distribuição almejada com relação aos n_i 's e considerando as Eq. (7.3) como vínculos, utilizamos dois multiplicadores de Lagrange α e β , e maximizamos o funcional

$$f(n_1, n_2, \dots, n_n) = \ln(W) + \alpha(\mathcal{N} - \sum_i n_i) + \beta(\mathcal{E} - \sum_i n_i E_i).$$

Note que esse procedimento é válido já que os máximos de W e de $\ln(W)$ ocorrem para os mesmos n_i , mas $\ln(W)$ é mais suave.

Tomando $\mathcal{N} \rightarrow \infty$ nesses cálculos, podemos aplicar a aproximação de Stirling, $\ln n! \cong n \ln n - n$, o que resulta em

$$n_i = e^{-\alpha - \beta E_i}.\tag{7.5}$$

Essa distribuição é chamada de canônica. Para determinar α e β é útil introduzir a função de partição canônica como uma soma sobre os estados microscópicos,

$$Z(\beta) = \sum_j e^{-\beta E_j}.\tag{7.6}$$

Logo, a partir das Eqs. (7.6) e (7.5) substituídas na Eq. (7.3) determinamos $e^\alpha = Z(\beta)/\mathcal{N}$.

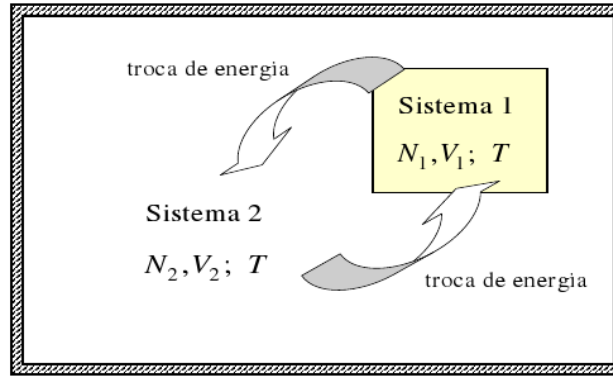


Figura 7.2: Esquema termodinâmico utilizado na dedução do ensemble canônico. O sistema 1 está acoplado energeticamente a um reservatório térmico, dito sistema 2, ambos estão à mesma temperatura de equilíbrio T e são isolados do resto do Universo.

Por fim, comparando os resultados aqui obtidos com fórmulas termodinâmicas [3], pode-se mostrar que $\beta = 1/k_B T$ e a energia livre de Helmholtz associa-se à função de partição canônica por meio de

$$F = -\ln Z(\beta)/\beta.$$

Consequentemente, a energia média $\langle E \rangle_{can}$ e o calor específico $\langle C_V \rangle_{can}$ no ensemble canônico são calculados através de

$$\begin{aligned} \langle E \rangle_{can} &= \frac{\mathcal{E}}{\mathcal{N}} = -\frac{\partial}{\partial \beta} \ln Z(\beta) \\ \langle C_V \rangle_{can} &= \left\langle \frac{\partial E}{\partial T} \right\rangle = \frac{1}{kT^2} \frac{\partial^2}{\partial \beta^2} \ln Z(\beta) = \frac{\langle E^2 \rangle - \langle E \rangle^2}{kT^2}. \end{aligned} \quad (7.7)$$

Dedução a partir do acoplamento com reservatório

Nesta seção procedemos à dedução do ensemble canônico partindo de argumentos termodinâmicos. O sistema de interesse é conectado a um reservatório térmico, como ilustrado na Figura (7.1), ambos trocam energia, mas permanecem isolados do restante do universo.

Utilizamos então a seguinte notação:

- S : é o sistema de interesse propriamente dito,
- S' : trata-se do reservatório térmico em que S reside, S é pequeno comparado a S' ,
- S^* : é o sistema conjunto composto por S e S' ,
- m : é a variável que indexa todos os estados de energia disponíveis para o sistema S ,
- E_m : é a energia do estado que corresponde ao índice m no sistema S ,
- E' : é a energia associada ao banho térmico,

- E^* : é a energia associada a S^* ,
- $\Omega'(E)$: denota o número de microestados disponíveis para uma energia particular E no reservatório térmico.

Assume-se que o sistema S e o reservatório S' estão em equilíbrio térmico à temperatura T . Nosso objetivo é calcular o conjunto de probabilidades p_m que caracteriza S no estado termodinâmico de energia E_m . Suponha para tal que S está em um estado indexado por m . Pelas definições acima a energia do sistema total S^* é dada por

$$E^* = E' + E_m. \quad (7.8)$$

E^* é constante, visto que S^* é um sistema isolado.

É crucial notarmos que a probabilidade p_m de encontrarmos S em um estado com energia E_m é proporcional ao número de microestados $\Omega'(E')$ disponíveis no reservatório S' . Portanto esta probabilidade é escrita como

$$p_m = C' \Omega'(E'), \quad (7.9)$$

onde C' é uma constante.

Logo, tomando logaritmos da Eq. (7.9) e reescrevendo-a, obtemos

$$\ln p_m = \ln C' + \ln \Omega'(E') = \ln C' + \ln \Omega'(E^* - E_m), \quad (7.10)$$

como E_m é pequena se comparada a E^* , uma expansão em série de Taylor pode ser executada neste logaritmo ao redor de E^* . Como uma boa aproximação, manteremos apenas os dois primeiros termos da expansão

$$\ln \Omega'(E') = \sum_{k=0}^{\infty} \frac{(E' - E^*)^k}{k!} \frac{d^k \ln \Omega'(E^*)}{dE'^k} \approx \ln \Omega'(E^*) - \frac{d}{dE'} \ln \Omega'(E^*) E_m. \quad (7.11)$$

Nesta expansão definimos a constante $\beta = \frac{d}{dE'} \ln \Omega'(E^*) = \frac{d}{dE'} \ln \Omega'(E') \Big|_{E'=E^*}$, associada termodinamicamente ao inverso da temperatura de equilíbrio do sistema.

Por fim, substituímos o resultado da Eq. (7.11) na Eq. (7.10), e obtemos

$$\ln p_m \approx \ln C' + \ln \Omega'(E^*) - \beta E_m,$$

ou

$$p_m = C' \Omega'(E^*) e^{-\beta E_m}. \quad (7.12)$$

Esta identidade torna-se exata se assumirmos tácitamente a hipótese de que E_m é pequena se comparada a E^* . O fator que antecede a exponencial na Eq. (7.12) pode ser interpretado como uma constante de normalização $C = C' \Omega'(E^*)$, assim

$$p_m = C e^{-\beta E_m}. \quad (7.13)$$

Impondo a condição de normalização,

$$\sum_m p_m = 1 = \sum_m C e^{-\beta E_m} = C \sum_m e^{-\beta E_m} \iff C = \frac{1}{\sum_m e^{-\beta E_m}} \equiv \frac{1}{Z(\beta)}, \quad (7.14)$$

obtemos $Z(\beta)$ que é conhecida como função de partição canônica.

Bibliografia

- [1] C. Cercignani, *Ludwig Boltzmann: the man who trusted atoms*, Oxford University Press, 1ª edição (1998).
- [2] D.H.E. Gross, *Microcanonical Thermodynamics: Phase Transitions in Small Systems*, Lecture Notes in Physics, vol. 66, World Scientific, Singapore, (2001).
- [3] D.A. McQuarrie, *Statistical Mechanics*, University Science Books, 2ª edição (2000).
- [4] C. N. Yang e T.D. Lee, Phys. Rev. **87**, 404 (1952); Phys. Rev. **87**, 410 (1952).
- [5] H.E. Stanley, *Introduction to phase transitions and critical phenomena*, Oxford University Press (1987);
G.B. Arfken e H.-J. Weber, *Mathematical methods for physicists*, Elsevier (2005);
www.mathworld.wolfram.com/
- [6] D.H.E. Gross, Phys. Rept. **279**, 119 (1997).
- [7] D.H.E. Gross, Phys. Chem. Chem. Phys. **4**, 863 (2002).
- [8] T. Padmanabhan, Phys. Rep. **188**, 285 (1990).
- [9] J. Barré, D. Mukamel e S. Ruffo, Phys. Rev. Lett. **87**, 030601 (2001).
- [10] J. Barré, S. Ruffo e N. Shreiber, Phys. Rev. Lett. **95**, 240604 (2005).
- [11] J. Barré, F. Bouchet, T. Dauxois e S. Ruffo, J. Stat. Phys. **119**, 677 (2005).
- [12] S. Ruffo, Eur. Phys. J. B **64**, 355 (2008).
- [13] J. Barré, D. Mukamel e S. Ruffo, *Dynamics and Thermodynamics of Systems with Long-Range Interactions*, Lecture Notes in Physics, Vol. **602**, 45, Springer (2002);
A. Campa, T. Dauxois, S. Ruffo, Phys. Rep. **480**, 57 (2009).
- [14] F. Staniscia, A. Turchi, D. Fanelli, P.H. Chavanis e G. De Ninno, Phys. Rev. Lett. **105**, 010601 (2010);
M. D'Agostino, F. Gulminelli, Ph. Chomaz, M. Bruno, F. Cannata, R. Bougault, F. Gramegna, I. Iori, N. Le Neindre, G.V. Margagliotti, A. Moroni e G. Vannini, Phys. Lett. B **473**, 21922 (2000);
H. Touchette, J. Stat. Mech. P05008, 1 (2010).

- [15] M. Blume, Phys. Rev. **141**, 517 (1966).
- [16] H.W. Capel, Physica (Amsterdam) **32**, 966 (1966); **33**, 295 (1967); **37**, 423 (1967).
- [17] A. Campa, S. Ruffo e H. Touchette, Physica A**385**, 233 (2007).
- [18] M. Costeniuc, R.S. Ellis, H. Touchette e B. Turkington, J. Stat. Phys. **119**, 1283 (2005).
- [19] M. Costeniuc, R.S. Ellis, H. Touchette e B. Turkington, Phys. Rev. E**73**, 026105 (2006).
- [20] M. Costeniuc, R.S. Ellis e H. Touchette, Phys. Rev. E**74**, 010105(R) (2006).
- [21] H. Touchette, M. Costeniuc, R.S. Ellis e B. Turkington, Physica A**365**, 132 (2006).
- [22] M. Costeniuc, R.S. Ellis, H. Touchette e B. Turkington, Prob. Geom. Integr. Syst. **55**, 131 (2007).
- [23] A.L. Lehninger, D.L. Nelson e M.M. Cox, *Lehninger principles of biochemistry*, W.H. Freeman, 4^a edição (2005).
- [24] C. Branden e J. Tooze, *Introduction to protein structure*, Garland Publishing (NY), 2^a edição (1999).
- [25] Yi Lu e S. Freeland, Gen. Biol. **7**, 102 (2006).
- [26] P.E. Leopold, M. Montal e J.N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992); J.D. Bryngelson, J.N. Onuchic, N.D. Socci e P.G. Wolynes, Proteins: Struct., Func. and Genetics **21**, 167 (1996).
- [27] http://www.cryst.bbk.ac.uk/PPS2/course/section7/os_pef.html
- [28] F. Eisenmenger, U.H.E. Hansmann, S. Hayryan e C.-K. Hu, Comp. Phys. Comm. **138**, 192 (2001).
- [29] F. Eisenmenger, U.H.E. Hansmann, S. Hayryan e C.-K. Hu, Comp. Phys. Comm. **174**, 422 (2006).
- [30] J.H. Meinke, S. Mohanty, F. Eisenmenger e U.H.E. Hansmann, Comp. Phys. Comm. **178**, 459 (2008).
- [31] A. Salam, *A unificação das forças fundamentais*, Jorge Zahar Editor, 1^a edição (1991).
- [32] C.M. Dobson, Nature: insight review articles **426**, 884 (2003).
- [33] R.W. Carrell e D.A. Lomas, The Lancet **350**, 134 (1997); D.J. Selkoe, Nature: insight review articles **426**, 884 (2003); F. Chiti e C.M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).

- [34] S. Ohnishi e K. Takano, *CMLS Cell. Mol. Life Sci.* **61**, 511 (2004);
- [35] C. A. Davie, *Brit. Med. Bull.* **86**, 109 (2008);
G.B. Irvine, O.M. El-Agnaf, G.M. Shankar e D.M. Walsh, *Mol. Med.* **14**, 451 (2008).
- [36] K. Blennow, M.J. de Leon e H. Zetterberg, *The Lancet* **368**, 387 (2006);
G.B. Irvine, O.M. El-Agnaf, G.M. Shankar e D.M. Walsh, *Mol. Med.* **14**, 451 (2008).
- [37] F.O. Walker, *The Lancet* **369**, 218 (2007);
R. Truant, R.S. Atwal, C. Desmond, L. Munsie e T. Tran, *FEBS Journal* **275**, 4252 (2008).
- [38] R.K. Obi e F.C. Nwanebu, *Afr. J. Clin. Exper. Microbiol.* **9**, 38 (2007).
- [39] D.R. Taylor e N.M. Hooper, *Mol. Membr. Biol.* **23**, 89 (2006).
- [40] I.H. Okamoto, *Rev. Neur.* **12**, 1 (2004).
- [41] S.B. Prusiner, Nobel Lecture: *Proc. Natl. Acad. Sci. USA* **95**, 13363 (1998);
K.-M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. Huang, R.J. Fletterick, F.E. Cohenu e S.B. Prusiner, *Proc. Natl. Acad. Sci. USA* **90**, 10962 (1993);
J. Shorter e S. Lindquist, *Nature Reviews: Genetics* **6**, 435 (2005);
K. Kuwata, N. Nishida, T. Matsumoto, Y.O. Kamatari, J.H.-Muto, K. Kodama, H.K. Nakamura, K. Kimura, M. Kawasaki, Y. Takakura, S. Shirabe, J. Takata, Y. Kataoka e S. Katamine, *Proc. Natl. Acad. Sci. USA* **104**, 11921 (2007);
P.M. Douglas, S. Treusch, H.-Yu Ren, R. Halfmann, M.L. Duennwald, S. Lindquist e D.M. Cyr, *Proc. Natl. Acad. Sci. USA* **105**, 7206 (2008).
- [42] F. Crick, *Nature* **227**, 562 (1970).
- [43] C. Junghans, M. Bachmann e W. Janke, *Phys. Rev. Lett.* **97**, 218103 (2006).
- [44] S. Schnabel, M. Bachmann e W. Janke, *Phys. Rev. Lett.* **98**, 048103 (2007).
- [45] J. Hernández-Rojas, J. M. G. Llorente, *Phys. Rev. Lett.* **100**, 258104 (2008).
- [46] M. Bachmann, H. Arkin e W. Janke, *Phys. Rev.* **E71**, 031906 (2005).
- [47] T. Neuhaus e J.S. Hager, *Phys. Rev.* **E74**, 036702 (2006).
- [48] T. Pawson e J. Schlessinger, *Curr. Biol.* **3**, 434 (1993);
F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley e E.I. Shakhnovich, *J. Mol. Biol.* **324**, 851 (2002).
- [49] R.S. Ellis, H. Touchette e B. Turkington, *Phys A* **335**, 518 (2004).
- [50] M.S.S. Challa e J.H. Hetherington, *Phys. Rev. Lett.* **60**, 77 (1988).

- [51] J.H. Hetherington, *J. Low Temp.* **66**, 145 (1987).
- [52] J.H. Hetherington e D.R. Stump, *Phys. Rev. D* **35**, 1972 (1987).
- [53] D.R. Stump e J.H. Hetherington, *Phys. Lett. B* **188**, 359 (1987).
- [54] M.S.S. Challa e J.H. Hetherington, *Phys. Rev. A* **38**, 6324 (1988).
- [55] J. Lukkarinen, *J. Phys. A: Math. Gen.* **32**, 287 (1999).
- [56] R.S. Johal, A. Planes e E. Vives, *Phys. Rev. E* **68**, 056113 (2003).
- [57] T. Morishita e M. Mikami, *J. Chem. Phys.* **127**, 034104 (2007).
- [58] M.S.S. Challa, D.P. Landau e K. Binder, *Phase Transitions*, **24-26**, 343, Gordon and Breavh Science Publishers (1990).
- [59] F. Bouchet e J. Barré, *J. Stat. Phys.* **118**, 1073 (2005).
- [60] R.S. Ellis, K. Haven e B. Turkington, *J. Stat. Phys. Phys.* **101**, 999 (2000).
- [61] B.A. Berg, *Fields Inst. Comm.* **26**, 1 (2000).
- [62] B.A. Berg, *Comp. Phys. Comm.* **153**, 397 (2003).
- [63] D.P. Landau e K. Binder, *A guide to Monte Carlo simulations in statistical physics*, Cambridge Univesity Press, 2^a edição (2005);
T. Tomé e M.J. de Oliveira, *Dinâmica Estocástica e Irreversibilidade*, EDUSP (2001);
S.A. Teukolsky, W.T. Vetterling e B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, W. H. Press 3^a edição (2007).
- [64] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller e E. Teller, *Jour. of Chem. Phys.* **21** (6), 1087 (1953).
- [65] A. Irback, C. Peterson, F. Potthast e O. Sommelius, *J. Chem. Phys.* **107**, 273 (1997).
- [66] A. Irback, C. Peterson e F. Potthast, *Phys. Rev. E* **55**, 860 (1997).
- [67] V. Tozzini, *Curr. Opin. Struc. Biol.* **15**, 144 (2005).
- [68] F.H. Stillinger, T.H.-Gordon e C.L. Hirshfeld, *Phys. Rev. E* **48**, 1469 (1993).
- [69] F.H. Stillinger e T.H.-Gordon, *Phys. Rev. E* **52**, 2872 (1995).
- [70] H.-P. Hsu, V. Mehra e P. Grassberger, *Phys. Rev. E* **68**, 037703 (2003).
- [71] S. Brown, N.J. Fawzi e T.H.-Gordon, *Proc. Natl. Acad. Sci. USA* **100**, 10712 (2001);
T. Vogel, T. Neuhaus, M. Bachmann e W. Janke, *Phys. Rev. E* **80**, 011802 (2009).

- [72] C. J. van Oss, D.R. Absolorn e A.W. Neumann, *Colloid & Polymer Sci.* **258**, 424 (1980).
- [73] K.M. Lundberg, C.J. Stenland, F.E. Cohen, S.B. Prusiner e G.L. Millhauser, *Chem. Biol.* **4**, 345(1997);
Y.S. Kim, T.W. Randolph, F.J. Stevens e J.F. Carpenste, *J. Biol. Chem.* **277**, 27240 (2002).
- [74] E.J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C.M. Dobson e C.V. Robinson, *Biophys. J.* **79**, 1053 (2000).
- [75] M. Kastner e M. Promberger, *J. Stat. Phys.* **103**, 893 (2001).
- [76] L. Casetti e M. Kastner, *Phys. Rev. Lett.* **97**, 100602 (2006).
- [77] J.V. Lopes, M.D. Costa, J.M.B.L. dos Santos e R. Toral, *Phys. Rev. E* **74**, 046702 (2006).
- [78] A. D. Sokal, *Monte Carlo methods in statistical mechanics: foundations and new algorithms*, Cargèse 1996, <http://citeseer.nj.nec.com/sokal96monte.html>.
- [79] M. Kac, G.E. Uhlenbeck e P.C. Hemmer, *J. Math. Phys.* **4**, 216 (1963).
- [80] M. Blume, V.J. Emery e R.B. Griffiths, *Phys. Rev. A* **4**, 1071 (1971).
- [81] D. Mukamel e M. Blume, *Phys. Rev. A* **10**, 610 (1974).
- [82] F. Leyvraz e S. Ruffo, *J. Phys. A.: Math. Gen.* **35**, 285 (2002).
- [83] R.B. Frigori, L.G. Rizzi e N.A. Alves, *Eur. Phys. Jour. B* **75**, 311 (2010);
R.B. Frigori, L.G. Rizzi e N.A. Alves, *Jour. of Phys. Conf. Ser.* **246**, 012018 (2010).
- [84] P. Westermark, C. Wernstedt, E. Wilander, D.W. Hayden, T.D. O'Brien e K.H. Johnson, *Proc. Nat. Acad. Sci. USA* **84**, 3881 (1987);
A.N. Roberts, B. Leighton, J.A. Todd, D. Cockburns, P.N. Schofield, R. Sutton, S. Holst, Y. Boyd, A.J. Day, E.A. Foot, A.C. Willis, K.B.M. Reid e G.J.S. Cooper, *Proc. Natl. Acad. Sci. USA* **86**, 9662 (1989);
E.T.A.S. Jaikaran e A. Clark, *Bioch. et Bioph. Acta* **1537**, 179 (2001).
- [85] <http://www.top500.org/>
- [86] T. Kennedy, *J. Statist. Phys.* **106**, 407 (2002).
- [87] S. Caracciolo, A. Pelissetto e A.D. Sokal, *Nucl. Phys. B (Proc. Suppl.)* **20**, 68 (1991).
- [88] R.B. Frigori, L.G. Rizzi e N.A. Alves, em preparação.

- [89] G.R. Smith e A.D. Bruce, *J. Phys. A: Mat. Gen.* **28**, 6623 (1995);
G.R. Smith e A. D. Bruce, *Europhys. Lett.* **34**, 91 (1996);
K.K. Bhattacharya e J.P. Sethna, *Phys. Rev E* **57**, 2553 (1998);
M. Nancias, C. Czaplewski e H.A. Scheraga, *J. Chem. Theory Comput.* **2**, 513 (2006).
- [90] T. Sugihara, J. Higo e H. Nakamura, *J. Phys. Soc. Jpn.* **78**, 074003 (2009).
- [91] P. Meier, N. Genoud, M. Prinz, M. Maissen, T. Rulicke, A. Zurbriggen, A.J. Raeber,
e A. Aguzzi, *Cell* **113**, 49 (2003);
C. Bate, S. Reid e A. Williams, *Jour. Biol. Chem.* **279**, 36405 (2004);
J. Masela, N. Genoudb e A. Aguzzib, *Jour. Mol. Biol.* **345**, 1243 (2005).