UNIVERSIDADE FEDERAL DE PERNAMBUCO
DEPARTAMENTO DE FÍSICA – CCEN
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA

TESE DE DOUTORADO

ASPECTOS ESPACIAIS E TEMPORAIS DO PROBLEMA DO ENOVELAMENTO PROTÉICO

por

Pedro Hugo de Figueirêdo

Tese apresentada ao Programa de Pós-Graduação em Física do Departamento de Física da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do título de Doutor em Física.

Banca Examinadora:

Prof. Sérgio Galvão Coutinho (Orientador-UFPE)
Prof. Edvaldo Nogueira Júnior (Co-orientador - UFBA)
Prof. Marcelo Albano Moret S. Gonçalves (Co-orientador - FVC)
Prof. Paulo Mascarello Bisch (IBCCF - UFRJ)
Prof. Jerson Lima Silva (IBM - UFRJ)
Prof. Marcelo Andrade de Filgueiras Gomes (DF-UFPE)
Profª. Rita Maria Zorzenon dos Santos (DF – UFPE)

Recife - PE, Brasil
Setembro – 2006

Figueirêdo, Pedro Hugo de
Aspectos espaciais e temporais do problema do enovelamento protéico / Pedro Hugo de Figueirêdo. – Recife : O autor, 2006.
xii, 124 folhas : il., fig., tab.

Tese (doutorado) – Universidade Federal de Pernambuco. CCEN. Física, 2006.

Inclui bibliografia e apêndice.

1. Mecânica estatística. 2. Enovelamento protéico. 3. Caminhantes aleatórios. 4. Séries temporais 5. Multifractais I. Título.

530.13 CDD (22.ed.) FQ2006-0019

Universidade Federal de Pernambuco
Departamento de Física – CCEN
Programa de Pós-Graduação em Física
Cidade Universitária - 50670-901 Recife PE Brasil
Fone (++ 55 81) 2126-8449/2126-8450 - Fax (++ 55 81) 3271-0359
http://www.df.ufpe.br/pg e-mail: posgrad@df.ufpe.br

Parecer da Banca Examinadora de Defesa de Tese de Doutorado

Pedro Hugo de Figueirêdo

ASPECTOS ESPACIAIS E TEMPORAIS DO PROBLEMA DO ENOVELAMENTO PROTÉICO

A Banca Examinadora composta pelos Professores Sérgio Galvão Coutinho (Presidente e Orientador), Marcelo Andrade de F. Gomes, Rita Maria Zorzenon dos Santos, todos da Universidade Federal de Pernambuco, Edvaldo Nogueira Júnior (Co-orientador), da Universidade Federal da Bahia, Marcelo Albano Moret S. Gonçalves (Co-orientador), da Fundação Visconde de Cairu, Paulo Mascarello Bisch, da Universidade Federal do Rio de Janeiro e Jerson Lima Silva, da Universidade Federal do Rio de Janeiro, consideram o candidato:

( ) Aprovado com Distinção (X) Aprovado ( ) Reprovado

Secretaria do Programa de Pós-Graduação em Física do Departamento de Física do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco aos quinze dias do mês de setembro de 2006.

Prof. Sergio Galvão Coutinho
Presidente e Orientador

Prof. Marcelo A. Moret S. Gonçalves
(Co-orientador)

Prof. Edvaldo Nogueira Júnior
(Co-orientador)

Profª. Rita Maria Zorzenon dos Santos

Prof. Marcelo Andrade de F. Gomes

Prof. Paulo Mascarello Bisch

Prof. Jerson Lima Silva

Em comunhão com aqueles que me ajudaram a desvencilhar os novelos da vida.

Agradecimentos

Certo dia ao longo de uma das crises acerca de nossos investimentos na carreira de físicos um amigo perguntou-me:

- Pedro, o que será que nos lembraríamos daqui a dez anos se abandonassemos a física ?

Ele me questionou se saberíamos por exemplo o que seriam bósons e férmions[1]. Terminamos a conversa rindo e dizendo que afirmaríamos que um era um vírus e outro era uma bactéria, mas não saberíamos identificar qual era o que. Hoje passados quase oito anos dessa conversa olho para trás e vejo um rastro formado por livros e artigos lidos, por listas de exercícios e exames cumpridos. Também é possível vislumbrar companheiros de farras, noites em praias e violões, comidas saborosas e discussões acaloradas. De certo é necessário que se registrem aqueles que contribuiram para que o estudante secundário que ingressou nessa instituição a dez anos atrás, se convertesse nesse candidato a doutor, e se eles não o tornaram um melhor físico, definitivamente foram decisivas em sua formação como um ser humano melhor.

Embora a enumeração possa ser perigosa, arrisco-me a citar alguns nomes mesmo que a memória me traia. Agradeço aos meus pais Maria e João, e a minha irmã Elisa que durante anos investiram paciência e apostaram em minha formação, mesmo sem entenderem muitas vezes o que eu fazia. Aos amigos do início da caminhada que faziam da biblioteca sua casa: Wilton, Paulo Henrique, Patrícia Façanha, Daniella Collier, Cibelle Nascimento, Chico Vieira e Mardson. Aos amigos de turma: Eric Parteli, Helinando Pequeno, Ana Maia, Clésio Leão, Frederico Brito e Jonas

[1]Se você não é físico não se preocupe, bósons e férmions são classificações técnicas para as partículas fundamentais constituintes da matéria. Bósons são partículas de spin inteiro enquanto que férmions são partículas de spin semi-inteiro.

Campelo. As garotas Karlla Adriana e Priscila Silva responsáveis por boa parte da alegria do ano de 2005. Aos matemáticos baianos Alex Ramos e Calitéia. Aos colegas "das turmas da frente" Laércio Dias, Mário Henrique, Mércia Liane, Fernando Parísio, Antônio de Pádua, Renê e Clécio Clemente; e "das turmas posteriores" Caio Veloso (encosto), Guga, Márcio Heráclito, Marcelo Alencar, Leonardo Cavalcanti, Felipe Fernando, Gerson Cortês. Ao meu afilhado e amigo de tantas escaladas: Ailton Fernandes, com quem compreendi que competência e humildade podem coexistir. Ao amigo José Ferraz, pelas tantas conversas e pensamentos compartilhados. À Cássia Donato, uma grande amiga, que de perto ou de longe sempre se soube fazer presente de forma positiva. Aos companheiros de grupo Hallan Silva, Washington Lima e Lenira e aos ex-companheiros Gustavo Camelo Neto e Alexandre Rosas.

Em especial a Marcelo Miranda e Josie Rabelo por terem colaborado para que o ano de 2006 fosse um ano de feliz aprendizado e cumplicidade no ingresso à vida adulta.

Aos mestres do ciclo básico Cláudio Furtado e Carlos Alberto. Aos primeiros mestres de pesquisa Marco Gameiro, Marcília Andrade Campos e Fernando Morais, pelos exemplos de ética, bom humor e profissionalismo.

Ao professor Sérgio Coutinho que ao longo destes seis anos de convivência mostrou-se um orientador atento, propiciando a liberdade, a confiança e a experiência necessárias para a consolidação desse trabalho. Aos professores Marcelo A. Moret e Edvaldo Nogueira, pelas colaborações e pelo competente aprendizado científico acerca dos sistemas protéicos.

Aos funcionários Joaquim, Ivo, Humberto, Cristina, Ricardo (in memorian), Paulo Pinto, João, Ana e Joana os quais deram e dão suporte para que a estrutura desse departamento funcione eficientemente.

Ao CNPq e a FACEPE que forneceram o suporte financeiro para esta pesquisa.

Por fim, acredito que daqui a dez anos bósons e férmions continuarão nos livros e poderão ser consultados, mas definitivamente as pessoas citadas acima continuarão em minha memória.

Resumo

A maneira na qual uma proteína se enovela a partir de uma espiral aleatória para um estado nativo *único*, em um intervalo de tempo relativamente curto, é um dos problemas fundamentais da biofísica molecular. É bem aceito que esta estrutura tridimensional única, característica de cada proteína e de sua seqüencia de amino ácidos, determina as funções da proteína. Nesta Tese, duas abordagens distintas serão empregadas para estudar aspectos gerais deste problema: 1) uma modelagem estocástica da cadeia principal e de formação de estruturas secundárias, para explorar os aspectos espaciais do estado nativo; 2) uma modelagem de dinâmica molecular, para analisar e caracterizar estatisticamente a evolução temporal da energia conformacional, durante o processo de enovelamento.

Na primeira abordagem, o modelo proposto gera uma cadeia principal com uma fração de estruturas secundárias através de um *caminhante aleatório angular* no espaço tridimensional, cuja trajetória, com passo de tamanho fixo e os ângulos diedrais (Φ e Ψ) das ligações peptídicas escolhidos por duas distribuições de probabilidades gaussianas, cujas médias estão associadas com as estruturas secundárias e a variância δ^2 como um parâmetro ajustável do modelo. Este modelo permite construir uma grande variedade de cadeias distintas, desde aquelas totalmente aleatórias às condizentes com dados experimentais. Algumas propriedades geométricas de proteínas globulares compostas por uma fração f de hélices-α e/ou fitas-β são particularmente estudadas. O comportamento de escala obtido para o raio de giração (R_g) em função do tamanho da cadeia (N); o grau de compactação (γ); a distribuição do número de coordenação (z_c) de carbonos C_α na estrutura e a energia total

envolvida, nestes contatos, foram explorados e comparados com dados extraídos de centenas de proteínas depositadas do *wwPDB* (*worldwide Protein Data Bank*). Os resultados encontrados mostram que, para a fração média $f \sim 0.6$ de estruturas secundárias (hélice-α e/ou fita-β), as cadeias geradas com distribuições de desvio padrão finito e próximo de $\delta \simeq 0.15\pi$ são mais compactas do que aquelas construídas com outros pares de ângulos diedrais. Independente dos detalhes dos mecanismos físico-químicos subjacentes, a construção de cadeias principais de proteínas com método geral proposto nesta Tese sugere que tais estruturas são governadas por distribuições de probabilidades estreitas e a estocasticidade desempenha um papel fundamental na sua compactação.

Na segunda abordagem, investigamos as propriedades multifractais de séries temporais da energia conformacional de pequenas estruturas em hélice-α, especificamente de uma família das polialaninas. Através do método de análise multifractal de flutuações destendenciadas (MF-DFA, do inglês *multifractal detrended fluctuation analysis*), estimamos o expoente de Hurst generalizado $h(q)$ e os associados expoentes de escala multifractal $\tau(q)$, para diversas séries, geradas numericamente por simulações de dinâmica molecular de sistemas em diferentes conformações iniciais. As simulações foram realizadas utilizando-se o campo de força GROMOS implementado no programa THOR. Os resultados mostram que todas as séries analisadas exibem um comportamento multifractal que depende do número de resíduos e da temperatura do sistema. Além disso, as propriedades multifractais das séries revelam aspectos importantes sobre a evolução temporal do sistema e sugerem que o processo de nucleação de estruturas secundárias, durante às visitas da proteína à sua hiper-superfície de energia potencial conformacional, são essenciais no processo de enovelamento.

Palavras-chave: Enovelamento Protéico, Caminhantes Aleatórios, Séries Temporais, Multifractais.

Abstract

The manner in which a protein folds from a random coil into a unique native state in a relatively short time is one of the fundamental puzzles of molecular biophysics. It is well accepted that a unique native three-dimensional structure, characteristic of each protein and determined by the sequence of its amino-acids, dictates protein functions. In this Thesis two distinct approaches are considered to study general aspects of such problem: 1) a stochastic modeling of the backbone chain of the protein secondary structures to explore the general spacial aspects of the native state; 2) an analysis of the time evolution of the protein conformational potential energy calculated during the folding process mimicked by methods of molecular dynamics.

In the first approach the proposed model generates a general backbone chain with a fixed fraction f of secondary like structures by means of a three-dimensional off-lattice random walk with fixed steps and the Φ and Ψ dihedral angles within the peptide bonds chosen by Gaussian probability distributions. Such probability distributions have their mean value corresponding to the angles associated with the chosen secondary structures and the variance δ^2 left as a free parameter to be determined. This model allows the construction of a great variety of backbone chains running from full random structures up to the biological ones observed in proteins. Some geometrical properties of globular structures, composed by a fraction f of α-helix and/or β-strands, were particulary studied. The scaling behavior of the ratio of gyration (R_g) with the chain size (N); the degree of compactness (γ); the distribution of coordination number (z_c) of the Carbon C_α atoms and energy

involved on such contacts were explored and compared with data of hundreds of proteins extracted from the *wwPDB* (*worldwide Protein Data Bank*). The results indicate that simulated structures are more compact when a fraction of $f \sim 0.6$ of secondary portions (α-helices and/or β-strands) are present than those built with other sets of dihedral angles, whenever the standard deviation of the probability distributions are finite and close to $\delta \sim 0.15\pi$. Independent of the details of all underlying physical chemistry mechanisms, building protein backbones with the method proposed in the present Thesis suggests that these structures are driven by narrow distributions leading to the conclusion that stochasticity has a fundamental role on the its compactness.

The second approach investigate the multifractal properties of the time-series of the conformational energy of small α-helix structures, in particular that of poly-alanine family. Using the multifractal detrended fluctuation analysis method (MF-DFA) the generalized Hurst exponent $h(q)$ and its associated multifractal scaling exponent $\tau(q)$ were estimated for several time-series numerically generated by molecular dynamic simulations considering distinct initial configurations. Such simulations were done using the force field GROMOS implemented by the software THOR. In general, the analyzed time series exhibit a multifractal behavior, which depends on the number of residues N and the temperature T of the system. Furthermore, whenever represented by the $h(q)$ or $\tau(q)$ spectra, the time-series multifractal properties reveal important aspects of the time evolution of the system. In particular, suggesting that the nucleation process of secondary structures, which should occurs during the *walk* of the protein on the corresponding portion of the conformational potential energy hyper-surface landscape, is essential for the folding process.

Keywords: Protein Folding, Random-Walks, Time Series, Multifractals.

Tese de Doutorado

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

"Não devemos nos sentir desencorajados pela dificuldade de interpretar a vida a partir das leis comuns da física."

Erwin Schrödinger - O que é vida ?

Mais notadamente nas duas últimas décadas, a física, a mais "natural" de todas as ciências e a de menor contato com o público leigo em geral, tem participado de um processo de interação com as demais áreas do conhecimento. Tal processo de intercâmbio que vai das ciências sociais aplicadas como a economia [1] até as ciências biológicas [2], passando pela imunologia [3, 4], pela psicofísica [5], pela sociologia [6], pela linguística [7] e pelo urbanismo [8], entre outras; tem permitido avanços não só na descrição e previsão dos fenômenos destas áreas, como também tem ajudado no desenvolvimento de uma série de ferramentas teóricas e experimentais que possiblitam o desbravamento de novos e velhos problemas físicos.

Com métodos comprovadamente eficazes desde os limites do mundo microscópico, como as dimensões dos motores moleculares [9] ao mundo regido pelas

escalas astronômicas [10], as contribuições da física mostram-se perceptíveis e robustas todas as vezes em que se faz necessário abordarmos os elementos constituintes de um sistema e suas interações, através de uma linguagem matemática que possa fornecer informações não só qualitativas como quantitativas acerca do comportamento macroscópico do sistema.

Neste contexto as colaborações ocorridas entre a biologia e a física tem sido um dos exemplos mais profícuos e emblemáticos. Contribuições historicamente impactantes da física à biologia remontam às suas próprias origens com as discussões de Galileu a respeito da altura característica dos seres humanos [11], passa pelas observações de Newton acerca da visão [12], chegando às sugestões de Erwin Schrödinger que abririam o caminho para a biologia molecular [13].

Do século XVI até os dias atuais [2] ambos os campos desenvolveram-se gerando uma grata interação que tem reunido, matemáticos, epidemiologistas, físicos, virologistas, cientistas da computação, químicos, neurologistas, teóricos evolucionários, dentre outras tantas áreas; todos em última análise movidos por perguntas fundamentais a cerca da vida como: Qual a sua origem? De que forma ela evoluiu até as formas que conhecemos hoje? Quais as possibilidades de sua existência sob outras condições distintas das observadas na Terra? De que maneira e em que nível dá-se o processamento de sinais neurológicos? Nesta tese abordaremos um dos aspectos centrais de um dos constituintes básicos de todas as formas de vida conhecida: as proteínas.

Diversas justificativas podem ser dadas na motivação de tal estudo. A primeira trata-se fundamentalmente do impacto que a determinação das estruturas protéicas tem para a indústria farmacêutica, uma vez que a ação dos chamados medicamentos inteligentes está reladionada à ancoragem dos agentes medicamentosos

às proteínas, como no caso dos inibidores de protease do HIV utilizados na terapia de tratamento para a AIDS. A segunda é que se credita à ocorrência de diversas doenças [14, 15] como anemias falciformes, fibroses, catarata, mal de Alzheimer e de Parkinson, bem como distúrbios como o da encefalopatia espongiforme ("mal da vaca louca"), à má-formação das proteínas ("miss-folding"), o que impediria sua correta função no organismo [16]. Desta forma a elucidação de tais doenças passa por uma profunda compreenssão do que ocorre estruturalmente com as proteínas em questão.

Além disso as analogias entre o problema biológico e seu mapeamento em problemas de cunho físico, bem como a utilização de ferramentas físicas tornam o assunto por si só interessante. É importante lembrar que nas duas últimas décadas a física, e em particular a denominada física da matéria condensada, responsável pelo estudo das propriedades estruturais e de transporte em sistemas físicos macroscópicos, tem dado grandes contribuições na caracterização dos chamados sistemas macios [17]. Exemplos de sistemas desse tipo são: os cristais líquidos; os polímeros; os colóides; e as emulsões, que embora descritos por interações distintas, são caracterizados por uma intensa resposta quando estimulados por campos externos. Da mesma forma, as proteínas como heteropolímeros flexíveis, são candidatas potenciais a serem investigadas pelos métodos físicos.

A primeira identificação das proteínas, como unidade fundamental biológica, é creditada ao químico Jöns Jacob Berzelius (1779-1848) (ver Figura 1). Quando nos seus estudos acerca da alimentação constatou que um óxido orgânico parecia ser básico para a nutrição animal, daí o nome originário do grego $\pi\rho\omega\tau\epsilon\iota o\xi$ que significaria primevo, primitivo. Posteriormente, no início do século XX, o químico alemão Emil Fischer (1825-1919) descobriu que as proteínas eram formadas por

cadeias peptídicas, compostas por aminoácidos. Desde então, muitos passaram a atribuir ao sucesso da vida no planeta Terra, o trabalho conjunto de dois grupos de macromoléculas: o ADN (ácido dessoxiribonucléico) e as proteínas. O primeiro responsável pelas instruções de construção e operação das estruturas e o segundo responsável pela construção *per se* [18], estabelecendo-se assim um esquema de "software" (programa) e "hardware" (máquina).

Figura 1.1: Jöns Jacob Berzelius propositor da existência das proteínas - 1838.

Em última análise seriam as proteínas as unidades responsáveis pelas principais atividades enzimáticas, como no caso da lactase responsável pela digestão da lactose; pelo transporte e armazenamento de substâncias, como no caso da hemoglobina; pela regulação e controle do sistema imunológico, tarefa dos anticorpos; pela contração dos músculos, ações da actina e miosina; pela transmissão dos impulsos nervosos no interior das células, como por exemplo, a rodopsina receptora dos bastonetes na retina; bem como no caráter estrutural dos organismos nos ligamentos e tendões constituídos por colágeno e queratina.

Para realizar adequadamente essas diferentes tarefas, acima citadas, as proteínas adotam uma única e bem definida configuração tridimensional, ditada por uma dada sequência de aminoáciodos, que recebe a denominação de estado enovelado ou nativo

[19]. A determinação e classificação dessas estruturas têm sido um dos grandes desafios aos pesquisadores da área, pois exigem a utilização de técnicas experimentais de boa resolução. Duas técnicas têm dado grandes contribuições neste sentido:a cristalografia de raios-X ou e a Ressonância Magnética Nuclear(RMN)[19]. Aplicações dessas técnicas tem possibilitado o aumento no número de proteínas catalogadas no WWPDB[1] (do inglês *Worldwide Protein Data Banking*). De fato esse número cresceu de 250 em 1988, para mais de 37392 na atualidade[2], graças sobretudo às contribuições oriundas de centros de pesquisa e universidades espalhadas em todo o mundo[3]. No entanto, hoje, um paradigma se apresenta em relação às pesquisas na área de biologia molecular, que pode ser resumidamente exposto da seguinte forma: como dar um passo além?, ou seja, como não apenas catalogar, mas, acima de tudo, prever a estrutura das proteínas e construí-las para um propósito específico?

Nesta Tese, apresentaremos duas abordagens para discutir a importante questão do enovelamento das proteínas. Na primeira, proporemos uma metodologia para construção de cadeias proteícas com um controlado processo de formação de estruturas secundárias. Na segunda, combinaremos recursos da dinâmica molecular e da análise estatística multifractal, para analisarmos propriedades estatísticas associadas às flutuações, presentes em séries temporais de energia potencial de proteínas. A Tese está organizda da seguinte maneira: No Capítulo 2, descreveremos as principais características dos sistemas protéicos, procurando relacioná-las ao problema do *enovelamento protéico* [20] e discutiremos os principais modelos propostos para abordagem desse problema. No Capítulo 3, apresentamos uma abordagem espacial simples, para a construção de polímeros que possuem várias das características

[1] http://www.wwpdb.org
[2] 27 de Junho de 2006
[3] Research Collaboratory for Strutctural Bioinformatics - http://www.rcsb.org

estruturais de proteínas reais. Diversas grandezas, tais como: o raio de giração, o número de contatos e a energia de interação entre os mesmos, serão obtidas e analisadas [21, 22]. No Capítulo 4, estudamos uma proposta para a caracterização multifractal das séries temporais, da energia conformacional de proteínas, obtidas por meio da dinâmica molecular de sistemas protéicos [23]. No Capítulo 5, apresentamos nossas conclusões gerais e perspectivas. No Apêndice A, derivamos resultados analíticos associados ao expoente de escala ν do modelo de Flory e no Apêndice B, apresentamos o sistema de coordenadas utilizado no modelo do caminhante Gaussiano.

Capítulo 2

Características Gerais dos Sistemas Protéicos

"... começar pelo princípio, como se esse princípio fosse a ponta sempre invisível de um fio mal enrolado que bastasse puxar e ir puxando até chegarmos à outra ponta, a do fim, e como se, entre a primeira e a segunda, tivéssemos tido nas mãos uma linha lisa e contínua em que não havia sido preciso desfazer nós nem desenredar estrangulamentos, coisa impossível de acontecer na vida dos novelos e, se uma outra frase de efeito é permitida, nos novelos da vida."

José Saramago - A caverna

2.1 Características gerais dos sistemas protéicos

A despeito de toda variedade de organismos vivos existentes, o processo da vida tal como conhecemos na Terra e como procuramos encontrar em outros pontos do Cosmo, baseia-se numa unidade química denominada aminoácido. Do ponto de

vista funcional cada aminoácido é um elemento com características próprias como: massa, carga, tamanho e hidrofobicidade que determinam as diferentes e possíveis estruturas denominadas proteínas. Essas "composições" de aminoácidos são determinantes na execução de funções que provêem a manutenção da grande estrutura que pode definir um organismo vivo.

Do ponto de vista químico um aminoácido é uma molécula constituída por um carbono central C_α conectado a um átomo de hidrogênio (H); um grupo amina (NH_2); um grupo ácido carboxílico ($COOH$) e um radical orgânico **R**, como ilustrado na Figura 2.1. A principal propriedade que determina as características estruturais dos aminoácidos e, portanto, das proteínas é a hidrofobicidade (apolaridade) ou hidrofilicidade (polaridade) do radical orgânico **R**, que está conectado ao carbono central denominado **carbono-C**$_\alpha$, o qual mantém, por sua vez, duas ligações simples: uma com o nitrogênio (**N**) do grupo amina e outra com o carbono **C**$_c$ do grupo carboxílico. Um **resíduo** é composto por um radical **R** e pelos átomos que fazem parte da cadeia principal da proteína. Finalmente devemos ressaltar que embora sejam conhecidos cerca de 100 aminoácidos, apenas 20 destes são geneticamente codificados, sendo portanto encontrados em todos os seres vivos.

A síntese protéica ocorre quando um resíduo conecta-se a outro por meio de uma **ligação peptídica**; tal ligação consiste de um condensamento do grupo carboxílico **C**$_c$ de um aminoácido ao grupo amina **N** do aminoácido seguinte, liberando assim uma molécula de água (H_2O). A repetição de tal processo por meio de sucessivas ligações peptídicas cria uma estrutura denominada cadeia principal ou "esqueleto" ("*backbone*") da proteína. Na Figura 2.2 apresentamos os vinte tipos de aminoácidos existentes, enquanto que na Tabela 2.1 listamos sua nomenclatura, separados em três grupos: o primeiro constituído por aminoácidos hidrofílicos (po-

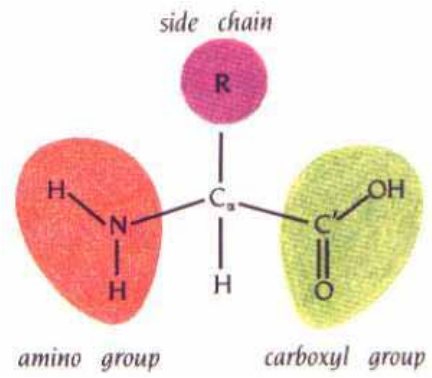

Figura 2.1: Estrutura química dos aminoácidos.

lares), o segundo por grupos eletricamente carregados e o terceiro por aminoácidos hidrofóbicos (apolares)[1].

Os aminoácidos formados pelos três grupos, acima citados, são moléculas quirais, ou seja, podem existir em duas diferentes formas, denominadas levógira (forma-**L**) e dextrógira (forma-**D**), em que uma é a imagem especular da outra. Na natureza apenas as formas levógiras são encontradas. Este fato é um dos grandes enigmas[2] que ainda permeiam a química e que parece possuir conexão estreita com a origem da vida em nosso planeta [18, 24]. Uma vez que o funcionamento dos sistemas biológicos está conectado com a especificidade estrutural das proteínas a existência dos dois tipos **L** e **D** poderia acarretar em deficiências no organismo.

2.1.1 Ligações químicas fundamentais

A princípio todos os potenciais e, consequentemente, as forças envolvidas

[1]O aminoácido Glicina (G=Gly), omitido na tabela 2.1 é constituido por apenas um único átomo de hidrogênio e possui características especiais de forma que muitas vezes é considerado um quarto grupo.

[2]Do ponto de vista das forças que unem as moléculas não há distinção entre uma forma ou outra

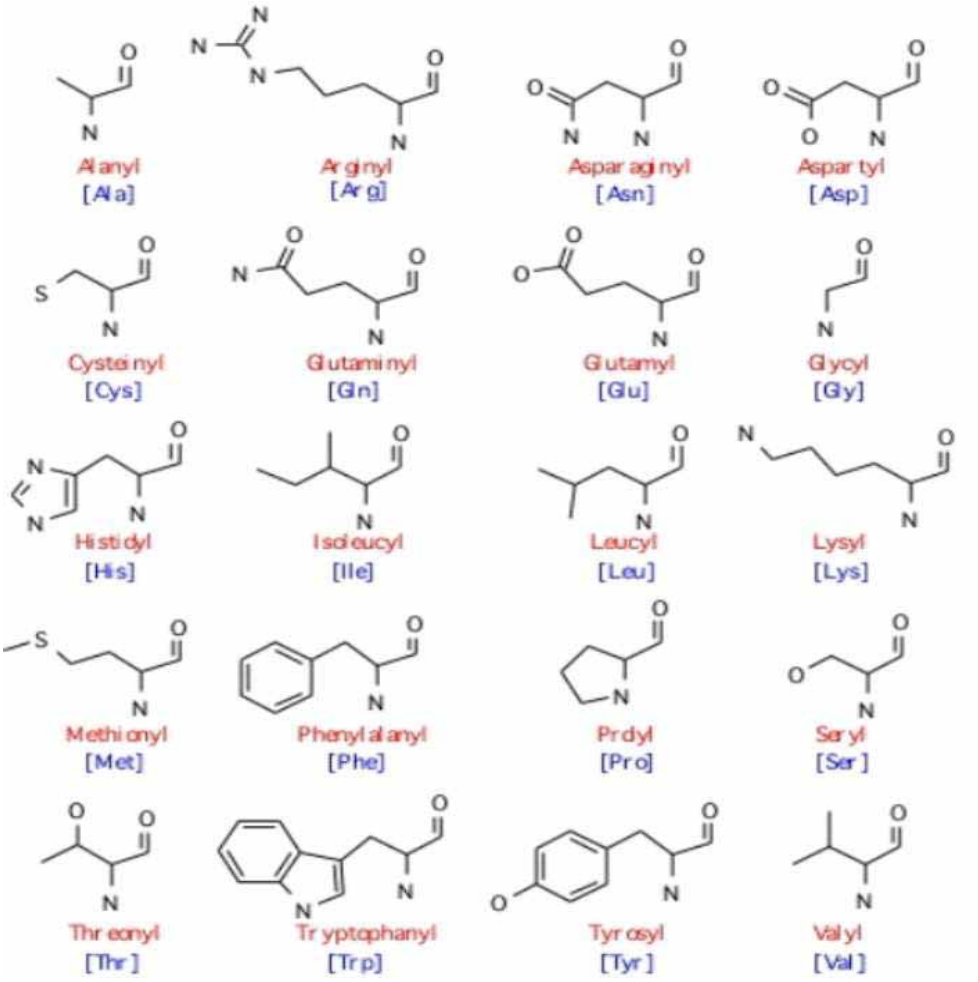

Figura 2.2: Relação dos 20 aminoácidos codificados pelos organismos vivos.

na estabilização conformacional das proteínas são de natureza eletrostática e encontram descrição quântica, de modo que o problema poderia, em princípio, ser solúvel exatamente. No entanto, esse procedimento torna-se impraticável do ponto de vista químico, uma vez que o número de resíduos constituintes de uma proteína pode alcançar a ordem de centenas, neste contexto falamos de ligações químicas efetivas. Para entendermos melhor essa questão; inicialmente vamos estabelecer a

Classificação	Nomenclatura
Hidrofílicos	Serina (S=Ser), Cisteína (C=Cys), Tirosina (Y=Tir), Asparagina (N=Asn) Glutamina (Q=Gln), Triptofano (W=Trp) e Teronina (T=Thr)
Hidrofóbicos	Alanina (A=Ala), Valina (V=Val), Fenilalanina (F=Phe), Prolina (P=Pro), Metionina (M=Met), Isoleucina (I=Ile) e Leucina (L=Leu)
Eletricamente carregados	Ácido Aspártico (D=Asp), Ácido Glutâmico (E=Glu), Lisina (K=Lis), Arginina (R=Arg) e Histidina (H=His)

Tabela 2.1: Classificação e nomenclatura dos 20 aminoácidos, sintetizados pelos organismos vivos, por hidrofobicidade, hidrofilicidade e carga elétrica.

escala de energia e as unidades utilizadas nesses sistemas. Frequente, trabalha com unidades de kilocaloria por mol (**kcal/mol**)[3]. Uma boa maneira de termos uma idéia da ordem de grandeza com estes valores é observarmos que o corpo humano a uma temperatura de 310K (37C°) corresponderia a uma energia de 0,642kcal/mol e que o banho térmico de uma sala à 298K (25C°) corresponderia a uma energia de 0,617kcal/mol, esta equivalência pode ser feita por meio da expressão k_BT onde k_B é a constante de Boltzmann e T é a temperatura absoluta do sistema. Estabelecido o padrão passaremos agora a examinar as forças [25, 26] e os tipos de ligações químicas envolvidas, bem como seus valores típicos. As ligações químicas e interações associadas a estabilidade estrutural das proteínas são:

- **Ligações covalentes**: Dentre as interações presentes na estabilização estrutural das proteínas as ligações covalentes são as mais intensas[4], com energias abrangendo variações entre 50 - 250 kcal/mol. Estas ligações são as responsáveis

[3]para converter em elétron-volts utilizar o fator $1eV = 23.06\,kcal/mol$

[4]para quebrar ligações dessa ordem por meio de radiação necessitaríamos incidir ondas eletromagnéticas na faixa de ultra-violeta ($\sim 10^{15}$Hz)

pela interação entre resíduos que não se encontram próximos na sequência dos aminoácidos, mas que se aproximam no estado enovelado, como no caso das pontes dissulfídicas que são formadas por dois átomos de enxofre presentes em diferentes partes da estrutura tridimensional. Tipicamente originadas de dois resíduos de Cisteínas, as pontes dissulfídicas são construídas a partir do processo de oxidação em que átomos de enxofre adjacentes "perdem" seus átomos de hidrogênio como descrito esquematicamente abaixo:

$$2(-CH_2SH) + 1/2O_2 \rightarrow -CH_2S - S - CH_2 + H_2O. \tag{2.1}$$

- **Efeitos Eletrostáticos:** Na Tabela 2.1 observamos que cinco dos vinte aminoácidos geneticamente codificados possuem carga líquida. Estas cargas estão expostas à ação de solventes, comumente a água, e de momentos de dipolo de outras moléculas, ou até mesmo de outros aminoácidos como, por exemplo, os hidrofílicos. Para se ter uma idéia dos valores típicos desses momentos de dipolo vale a pena observar que a molécula da água possui um momento de dipolo de 1.85D[5]. Outra fonte de carga líquida na estrutura protéica provém do desbalanceamento espacial de elétrons, nas ligações covalentes, este desequilíbrio tem sua origem nas diferentes eletronegatividades dos elementos que formam os aminoácidos. Os dipolos peptídicos são uma das grandes contribuições para a estabilidade local de macromoléculas biológicas. As energias associadas a estes efeitos eletrostáticos são descritos pelo potencial Coulombiano:

$$V_{ele} = \frac{1}{4\pi\epsilon_r} \sum_{i<j} \frac{q_i q_j}{r_{ij}}, \tag{2.2}$$

[5]Um próton de carga +e separado de um elétron de carga -e por uma distância de 1Å possui momento de dipolo de aproximadamente 4.8 Debye.

onde o somatório inclui as contribuições entre pares de átomos i e j, nas posições r_{ij} representa suas distâncias relativas, $q_i eq_j$ suas respectivas cargas e ϵ_r é a constante dielétrica do meio. Esta constante pode ter valores muito distintos dependendo do meio em que as cargas se encontrem. A água por exemplo, principal solvente biológico, possui uma constante dielétrica próxima a 80, enquanto que no interior de uma proteína esse valor cai para cerca de 4. Tal variação faz com que o potencial Coulombiano entre dois átomos separados por uma distância de 4Å tenha um valor de 1kcal/mol na água e de 20kcal/mol no interior de uma proteína. Em abordagens tipo dinâmica molecular para descrever o movimento de uma proteína [27], o valor da constante dielétrica desempenha fator decisivo e deve ser minuciosamente modelada afim de evitar divergências numéricas.

- **Ligações ponte de hidrogênio:** Como pode ser observado na estrutura do radical orgânico **R**, que determina os aminoácidos presentes na proteína, os átomos de hidrogênio **H** encontram-se ligados por meio de ligações covalentes a elementos fortemente eletronegativos como nitrogênio (**N**), oxigênio **O** e enxofre **S**. Devido a relação entre a eletronegatividade destes elementos e a eletropositividade do hidrogênio **H**, se estabelece um desbalanceamento espacial de carga que se traduz num momento de dipolo efetivo. A interação entre este momento de dipolo e um átomo parcialmente negativo em suas proximidades é denominada ligação de hidrogênio, ou ponte de hidrogênio. O alcance desta interação, ou seja, a distância característica entre um átomo doador e outro aceitador de carga é tipicamente da ordem de 3.5 Å e sua intensidade é da ordem de 1-7 kcal/mol.

- **Interações de van der Waals:** São exemplos típicos de ligações químicas

efetivas, compostas pela combinação de duas interações: uma primeira repulsiva de curto alcance e uma segunda atrativa de longo alcance. Do ponto de vista fundamental a primeira interação é de origem quântica, e resulta da repulsão das nuvens eletrônicas, devido ao Princípio da Exclusão de Pauli e a da força eletrostáticas entre os núcleos atômicos. A segunda interação resulta das flutuações quânticas do momento de dipolo dos átomos. Interações de van der Waals têm um comprimento característico [6] da ordem de 1.2 Å- 2.2 Å e um valor mínimo de energia em torno da energia térmica, ou seja, da ordem de décimos de kcal/mol e são descritas por:,

$$V_{vdw} = \sum_{i<j} \left[\frac{C_{12}(ij)}{{r^{12}}_{ij}} - \frac{C_6(ij)}{{r^6}_{ij}} \right], \tag{2.3}$$

onde r_{ij} representa a distância entre os átomos i e j dentro da estrutura e o somatório leva em conta todos os pares de átomos. Embora menos intensas que as interações discutidas anteriormente, as interações tipo van der Waals possuem um importante ingrediente de exclusão, que restringe o número de configurações acessíveis às proteínas[7]. Tal efeito pode se acentuar caso estejamos tratando de muitos contatos formados, o que ocorre quando superfícies moleculares complementares interagem, num mecanismo tipo chave-fechadura.

2.1.2 Formações estruturais típicas

Como discutido acima, as interações de van der Waals levam a uma restrição no número de possibilidades dos arranjos espaciais entre aminoácidos consecutivos. Aliado a este fato devemos observar que as ligações peptídicas são planares,

[6] Denominado raio de van der Waals.

[7] Este efeito de exclusão possui o nome técnico de efeito estérico.

de modo que rotações em torno dos grupos **N-C**$_\alpha$ e **C**$_\alpha$**-C**$_c$ de aminoácidos consecutivos, possuem valores específicos. Estas rotações são descritas por meio dos chamados ângulos diedrais ou de torção associados com a rotação do plano formado por 4 átomos (O,N,C,H), em relação à direção da cadeia principal da proteína. O ângulo diedral que corresponde à rotação em torno da ligação **N-C**$_\alpha$ é denominado Φ, enquanto que o correspondente à ligação **C**$_\alpha$**-C**$_c$ é denominado Ψ, conforme a Figura 2.3. A distribuição dos valores desses ângulos, como exposta na Figura 2.4, apresenta um caráter universal, na medida em que todas as proteínas conhecidas podem ser classificadas dentro de determinados grupos ou formas típicas. Primeiramente estabelecida pelo biofísico indiano G. N. Ramachandran, a distribuição dos valores de $\Phi e \Psi$, denominado *mapa de Ramachandran* [28, 29, 30] constitui-se hoje numa das caracterizações quantitativas mais importantes no estudo conformacional de proteínas.

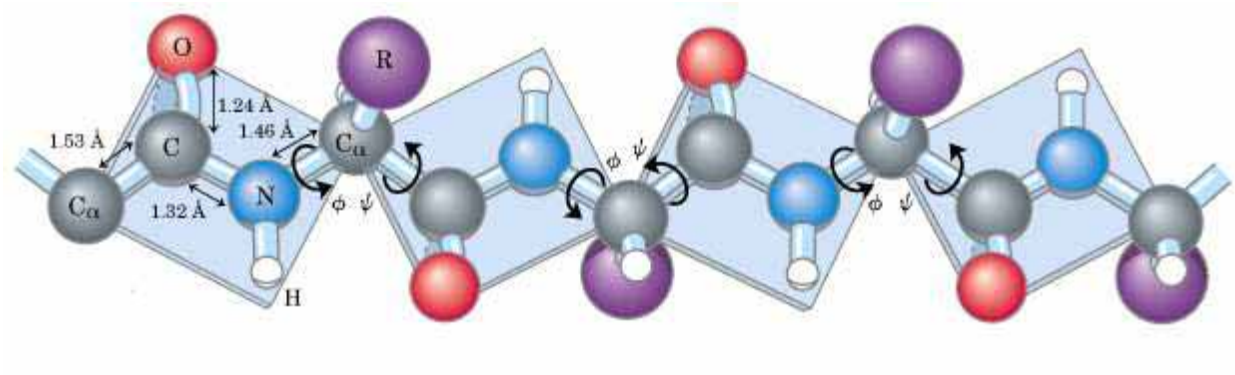

Figura 2.3: Valores dos ângulos de torção Φ e Ψ envolvidos nas ligações peptídicas.

Morfologicamente as proteínas podem ser classificadas como: **globulares** (estruturas esféricas) e **fibrosas** (morfologia tipo bastão, formada por várias cadeias proteícas). Nesta Tese trataremos basicamente de cadeias de proteínas globulares, cujo número de aminoácidos está entre 50 e 500 resíduos, o que corresponderia a uma

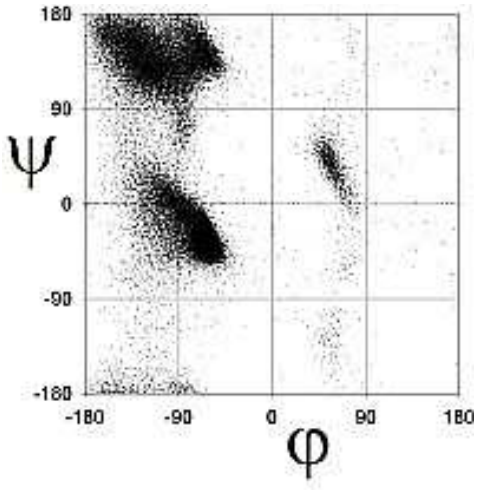

Figura 2.4: Mapa de Ramachandran exibindo regiões permitidas para os valores de Φ e Ψ nas estruturas protéicas

escala de comprimento que varia entre 25Å e 100Å. As proteínas globulares podem existir em quatro diferentes conformações: nativa (ordenada), "agregado globular", "pré-agregado globular" e desenovelada. Embora possamos observar uma distinção morfológica entre globulares e fibrosas em ambos, podemos identificar "motivos" ou níveis estruturais comuns a todas as proteínas:

1. **Estrutura Primária:** corresponde a sequência de aminoácidos que defines as proteínas. Proteínas podem ser comparadas com relação às suas sequências primárias, por uma técnica denominada de *homologia*, utlizada para determinar quão similar uma estrutura protéica é de outra. Similaridades entre polipeptídeos podem existir graças a permanência de estruturas selecionadas pela evolução biológica dos organismos.

2. **Estrutura Secundária:** devido à interação hidrofóbica entre a água, que envolve a proteína, e grupos que a compõem, certos grupos de aminoácidos agrupam-se de forma a criarem domínios os quais são caracterizados por

possuírem um núcleo hidrofóbico, uma superfície hidrofílica e curta dimensão espacial[8]. Estes padrões regulares são formados por ligações tipo ponte de hidrogênio entre a cadeia principal e os grupos **NH** e $\mathbf{C}_c\mathbf{O}$; correspondendo a valores específicos para o par de ângulos Φ e Ψ no mapa de Ramachandran. Estes arranjos espaciais recebem nomes particulares:

- hélice-α: estruturas clássicas da biologia molecular, são elas as formas preditas por Linus Pauling (em 1951) e, independentemente, pela dupla James Watson e Francis Crick (em 1953) como sendo os padrões energeticamente favoráveis à existência do DNA [31]. As hélices-α foram experimentalmente comprovadas, em 1958, por meio de técnicas de cristalografia de baixa resolução para a mioglobina, por Max Perutz e, posteriormente, com alta resolução por John Kendrew em 1958. De acordo com o mapa de Ramachandran, as hélices-α corresponderiam a região localizada no terceiro quadrante ($\Phi = -60°$ e $\Psi = -40°$) da Figura 2.4. Outros parâmetros que caracterizam as hélices são: seu comprimento, de aproximadamente 15 Å, seu número de resíduos, em torno de 3.6 (resíduos/passo) e o passo da hélice[9], que é da ordem de 5.4 Å. Como todos os átomos de hidrogênio dentro da hélices-α estão alinhados na mesma direção, os momentos de dipolo formados em cada aminoácido somam-se, de forma que toda a estrutura se comporta como um grande momento de dipolo. Devido a hidrofobicidade e a eletronegatividade característica de cada aminoácido, apenas alguns dos 20 listados na Tabela 2.1 são preferencialmente observados na formação de uma estrutura tipo $\alpha-$hélice; estando a alanina, o ácido glutâmico, a leucina e a metionina no grupo

[8] Quando comparados ao tamanho da proteína como um todo

[9] Tecnicamente a altura característica de uma volta (passo) é recebe o nome em inglês de *pitch*.

dos melhores formadores. A Figura 2.5 ilustra representações azimutais (c) e laterais (a,b,d) de uma estrutura secundária tipo $\alpha-$hélice.

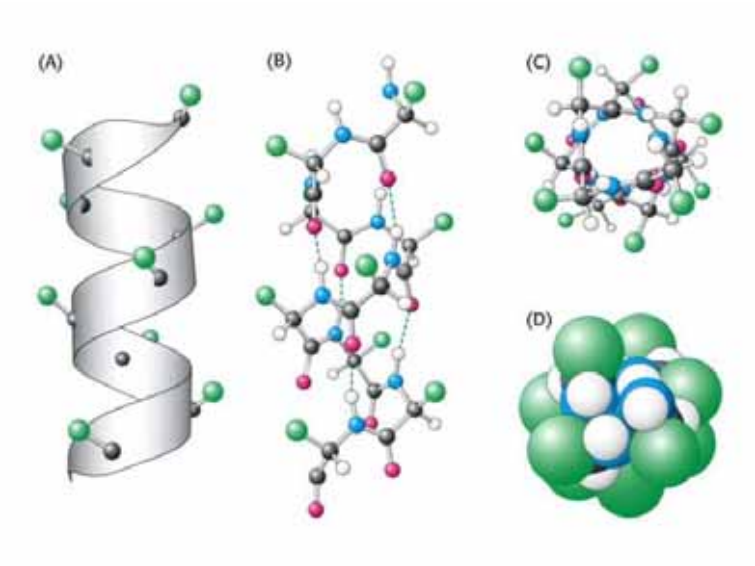

Figura 2.5: Diferentes representações de uma estrutura secundária em hélice-α

- fita-β: estruturas compostas de 5 a 10 resíduos e que são determinadas por ângulos presentes no segundo quadrante do mapa de Ramachandran tipicamente cujos valores estão entre $\Phi = -117°$ e $\Psi = 142°$. Comparativamente às hélices-α, os motivos tipo fita-β são bem menos compactos, apresentando-se com uma morfologia de fita, as quais podem se alinhar paralelamente ou anti-paralelamente, formando padrões, topologicamente similares a um plano, denominados folhas-β. Junto com as estruturas em hélice perfazem aproximadamente 60% do total de motivos tipicamente observados nas proteínas[10]. Na Figura 2.6 ilustramos esquematicamente, uma estrutura secundária tipo folha-β.

[10]Consultar http://en.wikipedia.org/wiki/Secondary_structure.

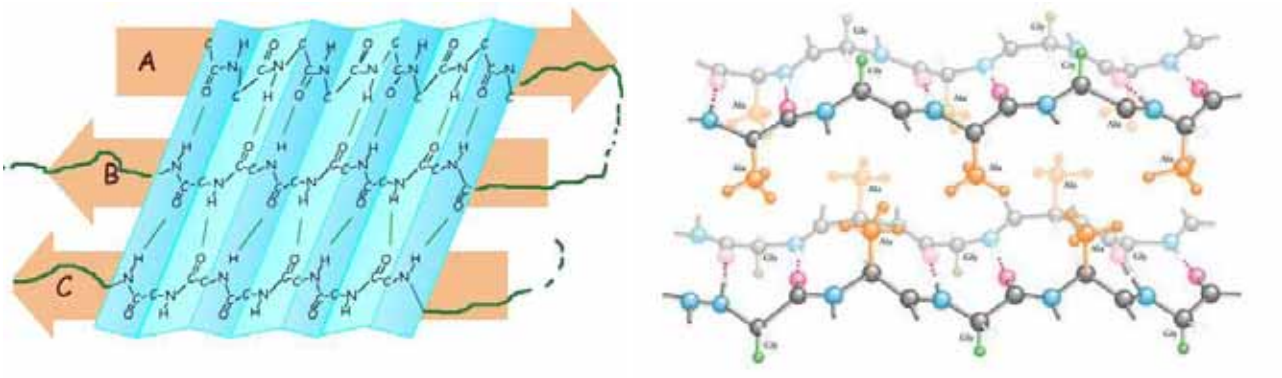

Figura 2.6: Estrutura secundária em folha-β e suas configurações paralela e anti-paralela

3. **Estrutura Terciária**: a unidade fundamental da estrutura terciária é um domínio, que pode ser definido como uma cadeia polipetídica ou parte da mesma, ou seja, essencialmente um domínio é uma estrutura secundária. Em geral, o conjunto formado pelos domínios globulares são estabilizados pelo empacotamento de motivos, hélice-α e/ou folha-β, ligados por pontes dissulfídicas. Estruturas terciárias possuem tipicamente 200 aminoácidos e são de crucial entendimento no processo de enovelamento, uma vez que cada domínio pode se empacotar separadamente. As conexões entre estruturas secundárias são feitas pelas denominadas regiões de *loop*, de formas irregulares e de tamanhos diversos. A Figura 2.7 mostra uma estrutura terciária esquemática, composta de hélices-α, fitas-β e *loops*.

4. **Estrutura Quaternária**: num nível hierárquico crescente, chegamos à composição de várias estruturas terciárias determinando o que é denominado de estrutura quaternária. Estes complexos globulares consistem de dois (dímero), três (trímero), quatro (tetrâmero) ou mais proteínas individuais; e podem ser classificados como homomérico ou heteromérico, caso sejam constituídos por

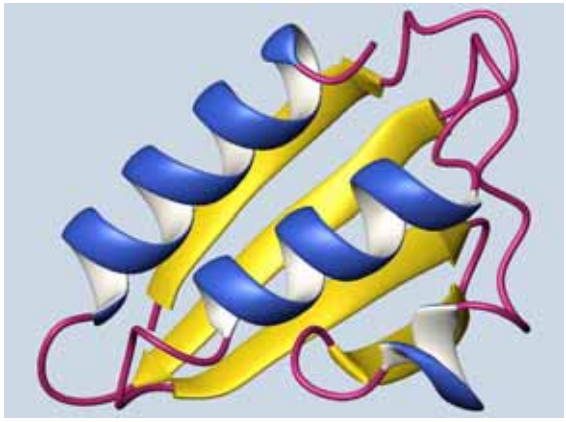

Figura 2.7: Estrutura terciária composta por hélices-α, folhas-β e loops

um único tipo ou mais tipos de cadeias de proteínas. Embora pareça estranho falar de uma proteína composta por proteínas, lembramos que esta classificação é esquemática, pois na realidade a proteína é a estrutura que desempenha uma função específica. Um caso clássico de estudo e bastante elucidativo é a hemoglobina, um hetero-tetrâmero composto por quatro "proteínas" distintas, duas hélices-α, duas folhas-β e seus *loops*, como mostrado esquematicamente na Figura 2.8.

Finalizando esta seção, destacamos ainda que à luz dos níveis estruturais, discutidos acima, vemos que o estado denominado "agregado globular", caracteriza-se pela ausência de uma cooperatividade da estrutura terciária da proteína, o que implica num raio hidrodinâmico, em média, 15% maior que o da estrutura nativa implicando num acréscimo de $\sim$ 50% em seu volume. Pelos mesmos argumentos o estado "pre-agregado globular" seria caracterizado por uma estrutura terciária "derretida" onde apenas aproximadamente 50% dos motivos secundários da estrutura nativa final estariam presentes. Observamos ainda que, existem indícios de que a chave para o entendimento da estrutura protéica encontra-se em sua estrutura

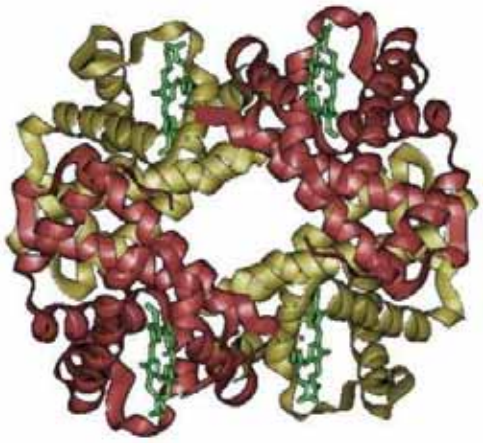

Figura 2.8: Representação de uma hemoglobina exibindo suas cadeias de proteínas constituintes.

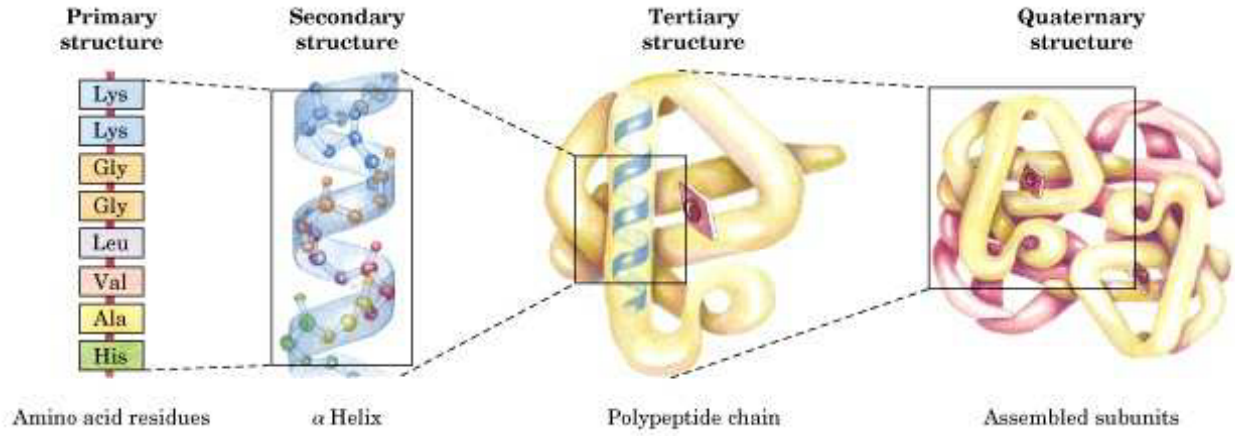

Figura 2.9: Diagramas esquemáticos das proteínas, exibindo os quatro níveis estruturais.

primária. Sob este aspecto cada proteína pode ser entendida como uma palavra construída a partir de um alfabeto de 20 letras (aminoácidos). Embora essa analogia pareça tentadora, é importante observar que cada palavra isoladamente não

traz informação relevante à semântica desse texto biológico, o que está intimamente ligada ao contexto (meio) em que se encontra [32, 33, 34]. Na Figura 2.9, resumimos a composição estrutural das proteínas mostrando sequencialmente sua hierarquia estrutural.

2.1.3 A hipersuperfície de energia e a hipótese termodinâmica

É importante observar que o padrão espacial da estrutura protéica é uma consequência direta das interações físicas de atração e repulsão entre as cargas que compõem a molécula e destas com as existentes no meio onde a mesma se encontra imersa. Neste processo cada átomo se compromete a satisfazer um princípio de minimização da energia da estrutura como um todo, mesmo que localmente se veja obrigado a estar submetido a um potencial mais elevado. Essa "frustração", típica de sistemas com muitas partículas, em outros problemas de interesse físico como nos vidros de spin [35, 36]. Nestes sistemas, os momentos magnéticos[11] com interações ferro e antiferromagnéticas, competem de modo a atingir uma configuração de menor energia, em contraposição às flutuações térmicas. Isto significa que, todos os átomos ou momentos magnéticos que os representam não podem localmente satisfazer uma configuração que minimize a energia com todos os elementos constituintes do sistema. No caso magnético o número de configurações estruturais possíveis de serem geradas cresce exponencialmente com o número de elementos envolvidos, de forma que, num processo de resfriamento, o sistema pode se encontrar preso em mínimos locais de energia do sistema. Estes estados, denominados meta-estáveis, encontram-se separados por barreiras de energia com altura infinita e entrelaçados em uma estrutura hierárquica.

[11] A origem de tais momentos magnéticos é puramente quântica.

Uma elucidativa analogia mecânica para tais tipos de problemas, consiste em imaginar a hipersuperfície de energia correspondente a cada uma das diversas configurações como sendo uma paisagem topográfica inundada formada por vales de diversas profundidades. Quando a estrutura encontra-se totalmente inundada a partícula não enxerga qualquer um dos vales. Contudo à medida que o nível da água baixa (correspondendo a uma diminuição da temperatura) a partícula pode ficar armadilhada num dos tantos vales que compõem a estrutura. Nesse quadro uma temperatura elevada pode ser identificada a um estado paramagnético do sistema magnético ou a uma molécula totalmente desenovelada. Enquanto que baixas temperaturas corresponderia a um estado ordenado ou ao "estado-nativo", ou enovelado da proteína.

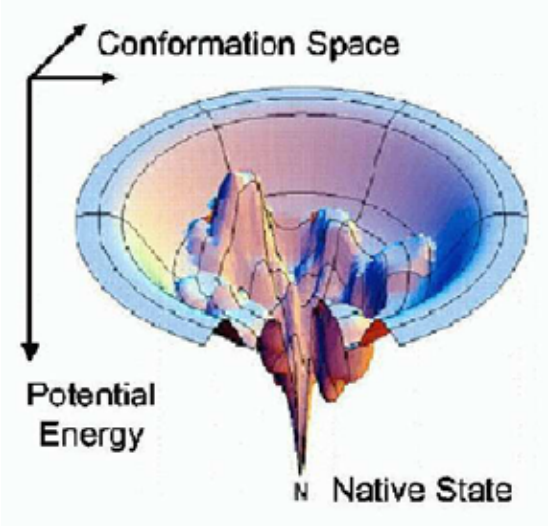

Figura 2.10: Representação da hipersuperfície de energia potencial característica das proteínas

Ao longo da Tese, utilizamos a palavra enovelada ou desenovelada referindo-se ao estado estrutural de uma proteína, mas sem indicarmos com precisão do que

se trataria. Tecnicamente define-se *enovelamento* como sendo o processo no qual uma proteína assume uma forma tridimensional específica, denominada também de *estado nativo*, que a permite realizar sua função biológica. Uma proteína, assim como outros tipos de heteropolímeros, podem enovelar-se ou desenovelar-se reversivelmente, por mudanças causadas pelo pH do meio aquoso, onde se encontre, ou por variações da temperatura desse meio. Neste caso, as proteínas usualmente perdem suas atividades biológicas e são denominadas *denaturadas*. Embora existam estudos, para o problema da denaturação, desenovelamento parcial ou total da proteínas, desde a década de trinta [37, 38, 39], foi somente a partir da década de cinquenta que os trabalhos de Anfisen [40] e outros levaram à formulação da chamada "hipótese termodinâmica", que estabeleceria bases quantitativas para a explicação do enovelamento.

Em linhas gerais, essa hipótese afirma que a informação contida na sequência de aminoácidos (estrutura primária) determinaria sua estrutura enovelada e que esta corresponderia a um mínimo global da energia livre configuracional do heteropolímero. Aparentemente, existem algumas exceções a esta regra, mas estas parecem ser atribuídos a estados meta-estáveis, armadilhados cineticamente durante o enovelamento [41]. Um caso intrigante que difere conceitualmente da referência [41] consiste de uma família de proteínas nativas intrinsecamente desenoveladas que não possuem propriedades estruturais uniformes [42].

Termodinamicamente as contribuições relevantes para caracterizarmos a separação energética entre os estados enovelado (nativo) e desenovelado (denaturado) de uma proteína são: a **Entalpia** e a **Entropia**. A primeira deriva das energias envolvidas nas ligações não-covalentes entre os peptídeos (essencialmente as interações hidrofóbicas, ligações de hidrogênio e iônicas). É sabido que as ligações

covalentes diferem muito pouco entre o estado nativo e denaturado, as ligações não-covalentes exibem um comportamento bem diverso, sendo muito intensas e de longo alcance no caso nativo e quase que irrelevantes no caso denaturado. Enquanto que a entropia conformacional possui valores baixos no caso nativo, uma vez que a organização estrutural é elevada, e muito elevados no caso denaturado, devido a desordem, também denominado "random-coil". É importante ressaltar que para que a entropia diminua na proteína o valor desta grandeza no meio deve se elevar.

A ordem de grandeza da barreira de energia que separa estes dois estados, denominada **energia livre**, é da ordem de 5 - 15kcal/mol, um valor típico de algumas pontes de hidrogênio. Contudo esta pequena diferença energética é um resultado envolvendo dois grandes números, comumente da ordem de centenas de kcal/mol. É importante notar que na ausência de algum fator que compense a entropia, o sistema tenderia a estar no estado denaturado, entropicamente mais favorável, assim a estabilidade da estrutura protéica, é assegurada por um ajuste químico, envolvendo variações de pH e temperatura do meio. Se do ponto de vista da previsão estrutural tal comportamento mostra-se tecnicamente complicado, é importante observar que isso assegura o sucesso da vida, uma vez que o funcionamento do organismo não deve envolver um elevado gasto energético para transformar proteínas de uma forma funcional para outra.

A questão que se coloca é a seguinte: como a proteína "encontraria" seu estado de mínima energia num tempo razoavelmente aceitável? Se partirmos da consideração, de que as proteínas realizam uma busca aleatória, entre as 3 regiões permitidas no mapa de Ramachandran (α, β e L) (uma estimativa bastante simplificada), até atingirem um estado "congelado", numa conformação de menor energia, então uma cadeia de 150 aminoácidos (uma proteína pequena) teria aproximada-

mente $3^{150} \simeq 10^{68}$ conformações distintas possíveis. Se o tempo requerido para converter uma conformação em outra for de $1ps = 10^{-12}$ s[12], então uma busca sistemática por todas as configurações levaria um tempo de aproximadamente 10^{56} segundos ou seja 10^{48} anos [13], conflitando com as observações experimentais, de que partindo de sua estrutura desenovelada, as proteínas mais rápidas atingem seu estado enovelado em tempos da ordem de 1ms e 1s *in vivo* e *in vitro*. Este comportamento aparentemente anacrônico denominado **Paradoxo de Levinthal** [43, 44, 45], em homenagem ao seu propositor o biofísico Cyrus Levinthal.

No final da década de 1960 Levinthal, mostrou que tal busca não seria possível, ou de outra forma, haveriam caminhos favorecidos nessa busca. As idéias iniciais para explicar o Paradoxo de Levinthal partem do pressuposto de que existem trajetórias específicas para o enovelamento, ou seja, haveriam moléculas que restringiriam o número de caminhos, assim as cadeias não necessitariam varrer todo o espaço configuracional [43]. A descoberta de estruturas intermediárias, parcialmente organizadas, como os agregados globulares e os domínios, ao longo do processo de enovelamento, parecem apoiar esta idéia e, embora ainda exista uma grande resistência entre os pesquisadores, alguns [46, 47, 48, 49] começam aceitar elucidações para o "paradoxo de Levinthal".

2.2 Abordagens para o problema do enovelamento protéico

As duas principais técnicas utilizadas para determinação da estrutura tri-

[12]Este seria o tempo em que uma ligação atômica se reorganizaria.

[13]Para efeito de comparação é importante notar que a idade do universo é estimada em 10-15 10^9 anos.

dimensional das proteínas são: a cristalografia de raios-X [50] e a Ressonância Magnética Nuclear (RMN) [51]. Aliadas a estas, outras como a espectroscopia de massa, a espectroscopia de infra-vemelho e a espectroscopia de ultra-violeta são fundamentais para caracterização dinâmica dos processos de enovelamento. Os primeiros experimentos, que inauguraram a forma de estudar o problema, do enovelamento foram realizados por Anfisen [40] seu principal objetivo nessas pesquisas era entender de que forma a conformação de equilíbrio de pequenas proteínas se alterava sob a variação cíclica de parâmetros como o pH e a temperatura do solvente, onde a mesma estava mergulhada. Como as proteínas atingiam a mesma configuração, sempre que as condições fisiológicas eram reproduzidas independente da trajetória, estes experimentos ajudaram a fundamentar a conjectura básica de que a conformação da proteína só dependeria da sequência dos aminoácidos que a constituiam.

Um exame do grau de enovelamento de uma proteína, ou seja, que fração de sua estrutura encontra-se enovelada pode ser medida através de dicroísmo circular [52]. Essencialmente o que estes experimentos descrevem é de que maneira a polarização da luz incidente numa dada amostra é afetada. Tal método se baseia na propriedade das estruturas secundárias hélice-α e fita-β girarem a polarização da luz[14]. Assim podemos acompanhar a população relativa de estruturas já enoveladas a medida que modificamos algum parâmetro externo, podendo então caracterizar a transição de fase estrutural dos heteropolímeros biológicos.

Um proteína típica consiste de algumas pontes salinas, uma centena de ligações de hidrogênio e milhares de interações de Van der Waals. Afim de simularmos este tipo sistema precisamos selecionar modelos apropriados que satisfaçam uma relação entre precisão e custo computacional. *A priori* poderíamos executar

[14] A onda eletromagnética é composta por campos elétrico e magnético, a direção de oscilação do campo elétrico é denominada de direção de polarização da luz.

cálculos com a precisão desejada a partir das posições e velocidades de todos os átomos que constituem a proteína, construindo um Hamiltoniano com todas as interações entre estas partículas, contudo tal abordagem computacionalmente não se mostra eficiente[15]. Primeiro, porque sabemos que algumas dessas interações são mais relevantes do que outras (a magnitude de determinadas interações pode chegar a ser de até duas ordens de grandeza maior do que outras); segundo, porque dependendo dos observáveis de interesse, um grau extremo de detalhamento pode ser totalmente irrelevante e por fim, pelo tempo e recursos computacionais que se dispõe. Assim, níveis de aproximação podem ser impostos aos sistemas de maneira a focalizarmos certos aspectos, como sua dinâmica temporal, sua geometria espacial ou aspectos da dinâmica evolucionária de um certo grupo de proteínas.

Diversos modelos têm sido propostos com diferentes metodologias e objetivos, a seguir descreveremos algumas classes desses modelos existentes, suas principais características, limitações e objetivos. A intenção aqui não é fornecer uma visão completa da "fauna" de modelos existente, mas sim discutir as principais idéias envolvidas em cada tipo de abordagem.

Nos **modelos de cinética química** diversas conformações são identificadas como alguns estados macroscópicos, comumente categorizados em três grupos. O primeiro grupo descreve os possíveis estados desenovelados (**U**, do inglês *unfolded*), correspondendo a todas as conformações não estruturadas da proteína. O segundo grupo consiste do estado nativo (**N**, do inglês *native*) da proteína que corresponderia a uma única conformação. O terceiro e último grupo caracteriza os estados intermediários (**I**, do inglês *intermediate*) que conectariam os estados **U** e **N**, por meio de uma cadeia de eventos descrita pela sequência:

[15]Neste caso estamos nos referindo à cálculos *ab initio*.

$$\mathbf{U} \rightarrow \mathbf{I}_1 \rightarrow \mathbf{I}_2 \rightarrow \ldots \rightarrow \mathbf{N}.$$

Cada estado encontra-se separado do outro por meio de uma barreira energética, de modo que as transições entre cada um dos estados são governadas por equações estocásticas, descritas pelas relações de Kramers [53, 54] utilizando-se de um perfil unidimensional de energia que dirige o caminho do enovelamento [55]. Este tipo de abordagem oferece uma descrição para as distriubições dos tempos de *folding*, podendo ser utilizado em conexão com potenciais de energia livre empíricos para predizer o efeito de mutações na estabilidade e na cinética de uma dada proteína [56]. Contudo, ele se mostra inadequado para descrever mecanismos moleculares na dinâmica de enovelamento. Além disso, ele descreve o sistema por meio de um parâmetro unidimensional, o perfil de energia, uma simplificação que não oferece nenhum aspecto geométrico da estrutura.

Uma segunda abordagem envolve uma compreensão mais detalhada da paisagem de energia livre e da entropia associadas à sua conformação espacial. Assim, nos denominados **modelos baseados em entropia**, são geradas um grande número de conformações através de simulações computacionais. As cadeias protéicas são então caracterizadas energeticamente por meio de um potencial de energia, obtido através de uma soma de termos de contato entre dois elementos da estrutura. Um exemplo clássico deste tipo de abordagem pode ser encontrado no modelo de Go [57], no qual cada aminoácido de uma proteína pode ser descrito através de uma representação onde cada átomo figura explicitamente na estrutura, ou por meio de uma representação simplificada em que cada aminoácido apresenta-se como uma conta esférica sem estrutura interna.

A idéia básica neste tipo de modelo, é que a geometria da proteína é o principal fator guiando o processo de enovelamento, e desta forma a energia li-

vre do sistema pode ser associada aos estados conformacionais por meio de um termo entrópico. Sua grande virtude é a capacidade de descrever a existência de fenômenos cooperativos na formação da estrutura espacial da proteína. Simulações realizadas para pequenas proteínas [58] confirmam, por exemplo, a existência de contatos específicos a partir dos quais a estrutura desenovelada atinge o estado nativo. Tal evidência indica que a proteína não se enovela de forma aleatória, seguindo na realidade uma sequência bem definida de passos. Por outro lado estes modelos não abordam a interação entre os 20 diferentes tipos de aminoácidos existentes na natureza e a possibilidade de que estes contatos possam formar interações inexistentes na estrutura nativa final. Ou seja, estes modelos negligenciam tanto o papel da sequência específica de aminoácidos para a formação de cada proteína, como a existência de estados metaestáveis que armadilhem o processo de enovelamento.

Uma outra abordagem termodinâmica para o problema do enovelamento, pode ser dada considerando-se a informação contida na sequência de aminoácidos que constitui a proteína. Nesta visão, a heterogeneidade das interações entre os aminoácidos produz uma paisagem de energia essencialmente rugosa, com diversos estados metaestáveis sendo formados. O objetivo dos **modelos baseados em energia** é entender de que forma uma proteína caracterizada por uma sequência específica de aminoácidos difere de um sistema aleatório. O principal elemento neste tipo de modelagem é a função de energia potencial que descreve a interação entre os contatos de dois aminoácidos da estrutura. Uma matriz para estas interações foi determinada por Miyazawa and Jernigan [59] em 1985.

O ponto de partida dos modelos baseados em energia é o estudo de cadeias formadas por uma sequência aleatória de aminoácidos, assim num modelo simplificado a energia de cada interação é escolhida aleatoriamente e a energia total para uma

configuração de N contatos é dada pela soma das energias de interação entre pares de aminoácidos na estrutura. Como consequência destas considerações estabelece-se um primeiro modelo denominado modelo de energia aleatória (REM do inglês, *Random Energy Model*) [60], o qual prevê um estado fundamental E_c e, a partir deste, um espectro contínuo. Como mostrado por Shakhnovich [61], a paisagem de energia para este modelo é composta por diversos mínimos locais, todos separados por pequenas barreiras de energia, que o distingue do comportamento de uma proteína real, a qual possui um mínimo global bem mais profundo do que os dos estados meta-estáveis. Nesse quadro, as características cinéticas do problema também tornam-se muito distintas daquela exibida por proteínas reais, uma vez que segundo esta prescrição uma proteína facilmente poderia ser armadilhada num desses estados meta-estáveis [62].

O desenvolvimento de computadores cada vez mais velozes tem permitido uma descrição mais detalhada da estrutura atômica das proteínas e dos potenciais de energia envolvidas na interação entre estes átomos. A construção da denominada **modelagem molecular**[27] baseia-se na simulação dos movimentos atômicos por meio de um campo de forças, para valores de temperatura e pressão conhecidos. É possível através desta abordagem analisar as interações provenientes entre a estrutura protéica e o solvente, onde a mesma se encontra imersa, tratando-o como um contínuo, ou até mesmo pela descrição explicita das moléculas que o compõem.

Essencialmente este tipo de modelo pretende resolver as equações de movimento de Newton através de um processo de integração numérica, o que permite investigar a evolução temporal entre várias conformações. Tecnicamente, a resolução das equações de movimento envolvem a utilização de potenciais efetivos clássicos parametrizados, os quais descrevem as interações intra e inter-moleculares.

O "coração" da dinâmica molecular reside em última instância, na função potencial escolhida, a qual deve ser realista o suficiente para fornecer descrições precisas da estrutura do sistema e ao mesmo tempo de fácil e rápida implementação numérica. Embora exista na literatura uma grande diversidade de potenciais e de programas, tanto de uso comercial quanto acadêmico (GROMACS, AMBER, CHARMM e THOR) [63, 64, 65, 66, 67], todos possuem dois fatores comuns: a existência de potenciais "ligados", usualmente representados por termos harmônicos, e de potenciais de interação à distância. Os primeiros são utilizados para descrever as ligações covalentes entre pares de átomos ângulos entre ligações químicas vizinhas e ângulos de torção em torno de ligações; enquanto que o segundo descreve interações entre átomos não ligados covalentemente, levando em conta as atrações e repulsões eletrostáticas, bem como as interações dipolares.

Como já discutido nas seções anteriores, devido ao elevado número de elementos constituintes, o número de graus de liberdade de um sistema molecular pode ser astronômico e desta forma torna-se impossível cobrir toda a superfície de energia. Como solução alternativa, os algoritmos tentam vasculhar não toda a superfície, mas sim o caminho que leva o sistema para uma configuração de menor energia. Para tanto entra em cena a segunda parte do problema de dinâmica molecular, a otimização de geometria. A otimização consiste essencialmente de um procedimento para obtenção, a cada passo de tempo fixado pelo algoritmo, do conjunto de cooordenadas que minimiza a energia do sistema como um todo.

Nesse espírito podem ser encontrados na literatura diversos tipos de algoritmos para efetuar tal minimização, tais como: o "steepest-descent" [68], o método dos gradientes conjugados [69] e os métodos de busca aleatória como o *"Generalized Simulated Annealing* [70]. É importante observar que assim como o potencial

de energia precisa ser acurado o suficiente para descrever as grandezas de interesse, a otimização precisa ser eficiente para que o sistema possa ser avaliado em tempos razoáveis. Dinâmicas que pretendem chegar a tempos da ordem de $1\mu s$ podem necessitar de tempos computacionais efetivos da ordem de dias, dependendo do tamanho do sistema, uma vez que cada passo da dinâmica gira em torno de 1fs. Embora se constitua de um poderoso método para a análise tanto de estruturas protéicas, como da interação destas com membranas e solventes, a grande limitação dos métodos de dinâmica molecular reside no fato de que atualmente a maior parte dos cálculos cobrem apenas uma pequena fração do tempo real envolvido no processo de enovelamento.

Um quarta abordagem são os **modelos de rede** indicados para estudar características gerais do enovelamento protéico. Tradicionalmente, se baseiam na aproximação de que a estrutura interna dos aminoácidos (seus átomos) pode ser negligenciada como consequência o caráter entrópico associado aos graus de liberdade internos podem ser também negligenciados. E assim, qualquer interação deverá ser considerada isotropicamente desta forma, os aminoácidos são representados por contas, conectadas em redes. Cada sítio pode estar vazio ou contendo uma conta, o que simularia o efeito do volume excluído. No caso de uma rede quadrada os ângulos de ligação entre cada aminoácido estão restritos aos valores $\pi/2$, π ou $3\pi/2$ radianos, enquanto que o comprimento das ligações tem um valor fixo caracterizado pelo parâmetro de rede. Estes modelos, também denominados de modelos minimalistas, estão preocupados em elucidar perguntas básicas acerca do enovelamento a saber: como uma cadeia polimérica pode enovelar-se num estado nativo único, como se dão os mecanismos de cooperatividade ao longo do processo, por quê algumas sequências primárias enovelam-se enquanto que outras não ? E o caráter da

frustração na evolução do processo [71].

O mais conhecido desta classe é o modelo HP [72], onde apenas dois tipos de aminoácidos são considerados: um denominado **H**, de caráter hidrofóbico e outro **P** de caráter polar estes aminoácidos são distribuídos aleatoriamente numa rede cúbica ou quadrada. É assumido nesse caso um potencial de interação entre sítios vizinhos dependendo dos aminoácidos que os ocupam (**H** ou **P**), dessa forma o modelo contempla a possibilidade de aminoácidos, que não são consecutivos sequência primária, poderem formar "contatos". Como observado por Tang [73] em sua revisão a respeito de modelos dessa natureza, a maior parte das estruturas geradas nesse tipo de modelo não se enovelam num único estado, o que não descreve a situação real protéica.

Em suas versões mais recentes [73], os modelos em rede consideram uma matriz de interação para um alfabeto composto por 20 aminoácidos, no mesmo contexto dos modelos baseados em energia. É da heterogeneidade destas iterações que se revela o caráter frustrado do sistema, que passa a exibir uma superfície de energia rugosa. Um dos grandes resultados obtidos com esse tipo de metologia é a a formação de estruturas elementares locais durante o processo de enovelamento. Essa formação ocorre numa sequência de eventos hierárquicos que se mostram fundamentais para o mecanismo de estabilização da estrutura como um todo [74].

Capítulo 3

Modelo de caminhantes angulares Gaussianos

"... você marcha, José! José, para onde? "

Carlos Drummond de Andrade - José

3.1 Caminhantes aleatórios e caminhantes auto-excludentes

Em 1828, o botânico escocês Robert Brown (1773-1858) realizou experiências nas quais observou, com a ajuda de um microscópio, que numa suspensão de grãos de pólen em água cada grão se movia irregularmente, numa trajetória errática que seria posteriormente denominada de Browniana [75]. Constatando que esse fenômeno era reproduzido com o uso de diversas substâncias orgânicas, Brown acreditou haver encontrado a molécula primitiva da matéria viva. Esse suposição foi posteriormente

refutada pelo próprio Brown, ao perceber que substâncias inorgânicas pulverizadas também possuiam comportamento similar. Após essa primeira incursão "equivocada" pela biologia, os *caminhantes aleatórios* (ou *"random walks" ; RW*), resurgiriam com fundamentação matemática, no contexto econômico em 1900, graças ao francês Louis Bachelier (1870-1946), em sua tese Teoria da Especulação[1], sobre opções de preço em mercados especulativos [76].

Uma descrição corpuscular da matéria, fundamentada num modelo de caminhante aleatório, foi introduzida por Albert Einstein (1879-1955) em 1905 [77]. Com a proposição do número de Avogadro, Einstein relacionou as grandezas estatísticas do movimento Browniano com o comportamento dos átomos e deu aos experimentalistas um método de contagem dos átomos, o que foi comprovado experimentalmente por Jean-Baptiste Perrin (1870-1942)[2] em 1908. Atualmente estudos de processos difusivos estão interessados em propriedades estatísticas associadas à caminhantes deslocando-se de maneira errante sobre um substrato. Essa dinâmica é recorrentemente utilizada como um conceito importante para se determinar as propriedades de escala de muitos fenômenos físicos [78, 79, 80, 81].

No modelo de caminhante aleatório, cada passo é completamente independente de todos os passos anteriores, tais processos estotásticos são também chamados de Markovianos. Contudo, para diversos processos físicos essa suposição não é apropriada. Ao idealizarmos um polímero, por exemplo, como uma longa cadeia flexível destituída de qualquer estrutura interna, onde cada ligação corresponde a um passo do caminhante, temos de levar em conta a exclusão espacial. Uma caminhada aleatória submetida a tal condição é denominada *caminhada aleatória*

[1] Orientada pelo também matemático Henri Poincaré(1854-1912), a tese de Bachelier foi preterida por seus contemporâneos e permaneceria na obscuridade até meados da década de 1960.

[2] Por suas medições Perrin receberia o prêmio Nobel em 1926.

auto-excludente (ou *self-avoiding walk; SAW*). Na Figura 3.1, apresentamos padrões típicos gerados por uma caminhada aleatória e uma caminhada auto-excludente, ambas sobre uma rede quadrada. Em geral, as quantidades geométricas analisadas nesse tipo de modelo são: o deslocamento quadrático médio em relação à origem $< r^2 >$ e o raio de giração da estrutura final R_g, definido como a distância quadrática média de cada posição ao centro de massa da estrutura,

$$R_g \equiv \sqrt{\sum_{i=1}^{N} \frac{\rho_i^2}{N+1}}, \tag{3.1}$$

onde ρ_i é a distância de cada um dos elementos da estrutura ao centro de massa e N é o número de passos do caminhante.

Tais grandezas exibem uma relação de escala, ou seja, o raio de giração, por exemplo, exibe um comportamento tipo lei de potência com o número de passos N do caminhante,

$$R_g \sim N^{\nu}, \tag{3.2}$$

onde ν é o expoente que caracteriza o tipo de difusão subjacente ao processo, com a seguinte classificação:

$$\begin{cases} \nu < 1/2 & \text{(subdifusivo)}, \\ \nu = 1/2 & \text{(difusivo)}, \\ \nu > 1/2 & \text{(superdifusivo)}. \end{cases} \tag{3.3}$$

Um modelo tipo campo médio para formação de estruturs poliméricas embebidas em um bom solvente, desenvolvido por Flory[82, 83], prevê um expoente

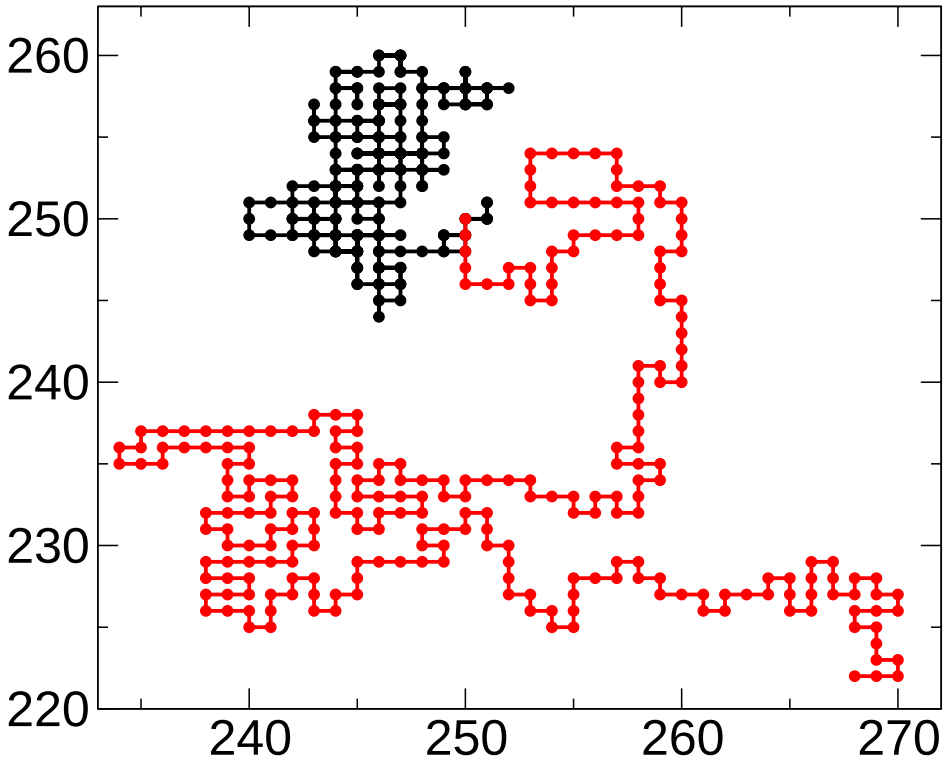

Figura 3.1: Padrões típicos, obtidos através de simulação numa rede quadrada, para caminhantes aleatórios (em preto) e caminhantes aleatórios auto-excludentes (em vermelho), ambos com 250 passos e partindo do ponto $(250, 250)$.

dependente da dimensão D, dado por:

$$\nu_{Flory} = \frac{3}{D+2}, \tag{3.4}$$

de modo que $\nu_{Flory} = 3/5$ para o caso em três dimensões pertencendo, também, a um regime superdifusivo, este resultado encontra-se reproduzido no Apêndice A desta Tese. Recentes simulações [84] de *RW* e *SAW* em três dimensões apresentam resultados diferentes para o expoente com $\nu_{RW} = 1/2$ e $\nu_{SAW} = 0.588$, indicando que no caso auto-excludente a partícula escapa mais rápido da origem que no regime

difusivo.

3.2 Modelo de caminhantes angulares Gaussianos

Ao tratarmos de cadeias protéicas, observamos que o raio das estruturas com o número de aminoácidos, também obedece um comportamento de escala, como pode ser observado no gráfico log-log experimental, baseado em 1826 cadeias protéicas, exibido na Figura 3.2. No contexto discutido na seção anterior, este comportamento pertence a um regime sud-difusivo, com $\nu_{exp} = 0.40 \pm 0.02$, diferente daquele previso por Flory, $\nu_{Flory} = 0.60$ e observado no contexto de superfícies amassadas como indicado por Gomes e colaboradores [85]. Constatamos dessa forma que mesmo as estimativas construídas em argumentos de campo médio, onde a entropia da superfície da estrutura é levada em conta, não é capaz de descrever esta característica geométrica básica.

Como discutido no capítulo 2, diversas abordagens para descrição das propriedades estruturais das proteínas têm sido propostas. Devido ao número astrônomico de possíveis configurações para proteínas globulares (compostas por 50 a 500 aminoácidos), metodologias convencionais fundamentadas em Monte Carlo ou dinâmica molecular tornam-se impraticáveis, devido ao seu alto custo computacional.

Nesta seção, apresentamos uma estratégia alternativa, baseada num modelo de caminhante aleatório em três dimensões, como forma de construirmos cadeias protéicas com diferentes comprimentos e porcentagens de estruturas secundárias. No modelo, cada passo possui um comprimento radial l_0 fixo, mas os ângulos diedrais, Φ e Ψ que compõem a cadeia são escolhidos independentemente por meio de uma distribuição de probabilidades Gaussiana, seguindo a proposta de Shaw e colaboradores [79]. Os valores médios e o desvio de cada distribuição são definidos de acordo

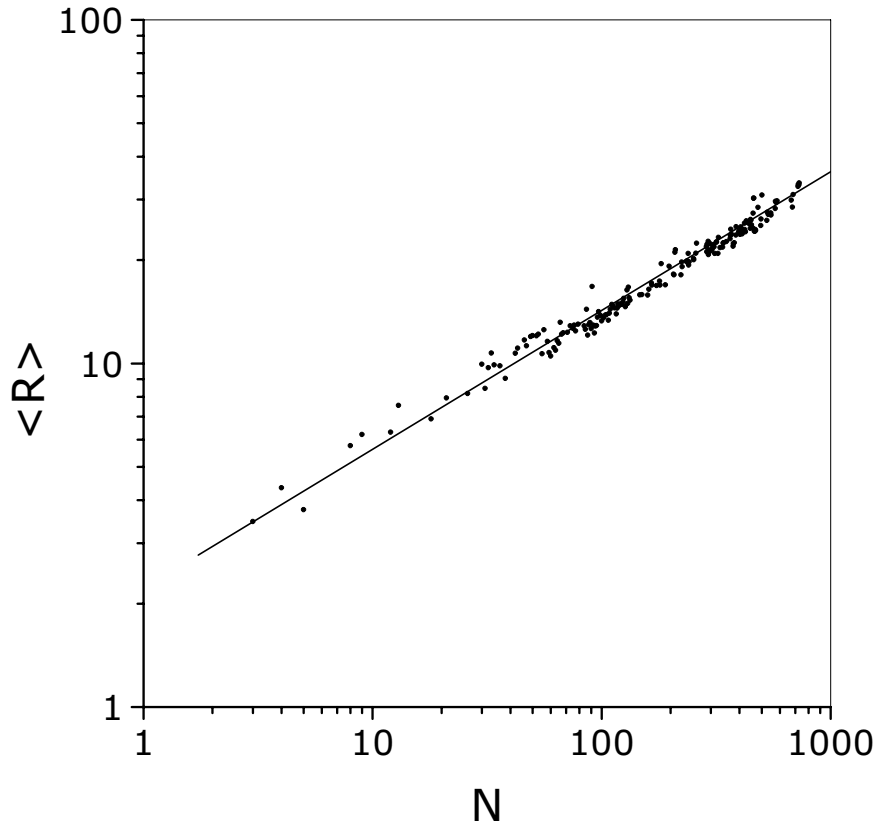

Figura 3.2: Comportamento do raio médio $< R >$ em função do número de aminoácidos N para um conjunto de 1826 cadeias protéicas, com expoente $\nu \approx 0.40 \pm 0.02$. A linha contínua indica o ajuste linear dos dados

com as regiões permitidas para Φ e Ψ presentes no mapa de Ramachandran. Aqui utilizamos os valores calculados por estatíticas de diversas estruturas secundárias, propostos pelo programa PRELUDE [86] a Tabela 3.1 indica sete possíveis pares de ângulos diedrais e suas conformações associadas.

Desta forma, as distribuições Gaussianas para os ângulos são explicitamente descritas como:

$$P(\Phi) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(\Phi - \Phi_o)^2}{2\delta^2}\right), \tag{3.5}$$

e

$$P(\Psi) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(\Psi - \Psi_o)^2}{2\delta^2}\right), \tag{3.6}$$

onde δ é o desvio padrão de cada distribuição, enquanto que Φ_o e Ψ_o são respectivamente os valores médios dos ângulos diedrais Φ e Ψ.

Φ_o	Ψ_o	Conformação
-65°	-40°	A
-89°	-1°	C
-117°	142°	B
-69°	140°	P
78°	20°	G
103°	-176°	E
-83°	133°	O

Tabela 3.1: Sete possíveis pares para ângulos diedrais (Φ,Ψ) e suas conformações associadas [86]. Configurações em hélice-α denotadas por A e folha-β por B.

Para simularmos proteínas com uma determinada porcentagem de estruturas secundárias f, fixamos o número de passos do caminhante N e estabelecemos um processo de crescimento descrito pelas seguintes regras:

1. Nos primeiros $l_g \times f$ passos escolhe-se **um** dos pares de ângulos da Tabela 3.1 e por meio das distribuições descritas nas Equações 3.5 e 3.6, construímos sequencialmente a próxima posição como função da posição anterior.

2. Nos próximos $l_g \times (1 - f)$ passos utilizamos, a cada passo, ângulos aleatoriamente escolhidos entre os sete possíveis pares, utilizando mais uma vez as distribuições descritas nas Equações 3.5 e 3.6, construímos sequencialmente a próxima posição como função da posição anterior.

3. Nos próximos l_g passos as regras 1 e 2 são repetidas, até construirmos uma cadeia de tamanho N.

A determinação da posição (x_i, y_i, z_i), como função da posição anterior $(x_{i-1}, y_{i-1}, z_{i-1})$, envolve uma transformação entre coordenadas cartesianas e internas, descritas pelos ângulos Φ e Ψ, como observado por Park e colaboradores [87] [3]. Neste modelo minimalista, a construção do esqueleto peptídico envolve apenas os ângulos diedrais. Todos os potenciais efetivos descritos no Capítulo 3, sejam eles de contato ou de ação à distância, são levados em consideração indiretamente através da escolha dos ângulos da Tabela 3.1. Uma vez que a distribuição de ângulos diedrais encontra-se confinada às regiões permitidas pelo mapa de Ramachandran, espera-se que os fenômenos estéricos estejam inclusos nessa abordagem implicitamente. Devido à sua simplicidade computacional, o método permite a construção de uma elevada quantidade de amostras, ou seja, de diferentes conformações (da ordem de 10^4).

É importante perceber, que nesse modelo, a **regra 1** induz um crescimento ordenado, durante um comprimento característico $(l_g \times f)$, enquanto que a **regra 2** descreve a quebra desse padrão estrutural, quebra esta que pode ser entendida como resultante da instabilidade das forças envolvidas na formação de conformações específicas, como no caso dos momentos de dipolo efetivos das estruturas hélice-α.

Pode-se então simular diversas destas configurações, variando-se a fração f e o tipo de estrutura secundária utilizada na regra 1 do modelo. Em nosso estudo utiliz-

[3] Para detalhes consultar o Apêndice B desta Tese.

amos: (a) estruturas hélice-α; (b) folhas-β ou (c) uma mistura de hélice-α e folha-β. A escolha desses motivos estruturais decorre de sua relevância nos estágios iniciais do enovelamento protéico e por estes serem os blocos fundamentais na formação de estruturas terciárias [88, 89]. Nas Figuras 3.3, 3.4 e 3.5 exibimos configurações típicas compostas por $N = 250$ resíduos, para os três casos acima especificados, assim como os respectivos os mapas de Ramachandran para um conjunto de 100 amostras, em cada um desses casos. Em todas as simulações a largura da distribuição vale $\delta/\pi = 0.1$, o comprimento $l_g = 100$ e o comprimento característico entre cada resíduo (distância radial) vale $l_o = 3.8$ Å.

3.3 Análise das grandezas relevantes

Uma vez criadas diferentes configurações, através do algorítmo descrito na seção anterior, utilizamos algumas grandezas a fim de caracterizar estruturalmente as cadeias geradas artificialmente e compará-las às cadeias reais. Seguindo a sugestão de Tang e colaboradores [90], elegemos o raio de giração R_g; o comprimento de contorno l_c; o número de contatos n_c; o número de coordenação z_c e a energia entre os contatos E. Além destes parâmetros introduzimos outra quantidade denominada "parâmetro de compactação", definido como:

$$\gamma \equiv \frac{R_g}{D_{max}}, \tag{3.7}$$

onde R_g é o raio de giração da estrutura e D_{max} é a maior distância entre dois resíduos da mesma. Discutiremos a seguir a relevância de cada uma dessas grandezas e de que forma os parâmetros de nosso modelo as descrevem.

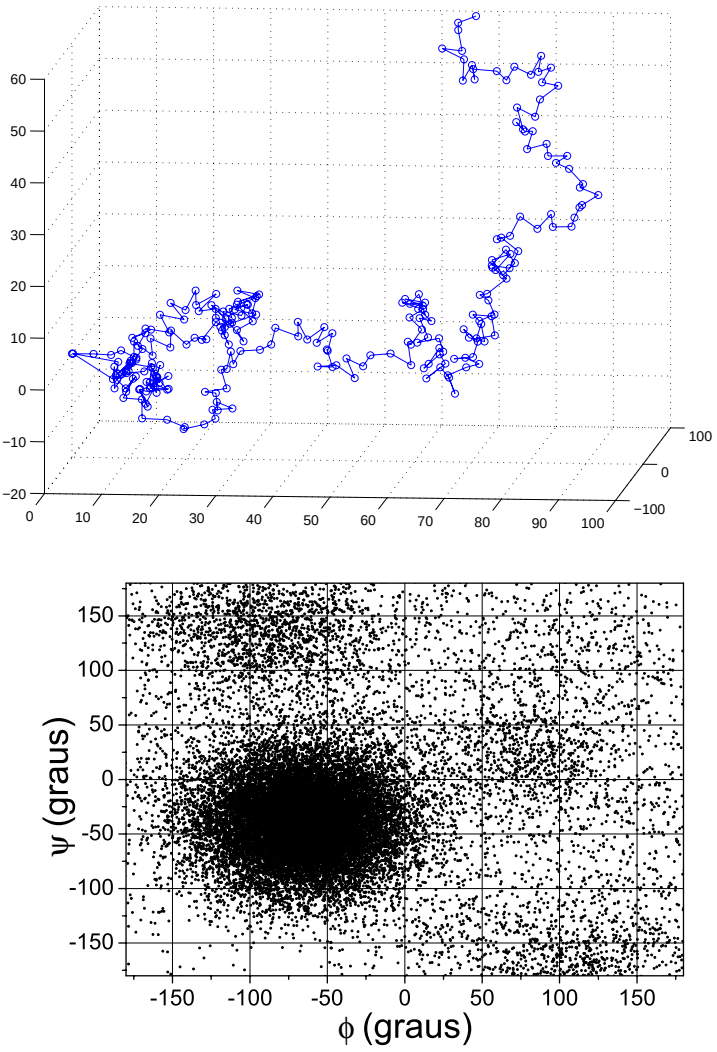

Figura 3.3: (a) Padrão típico de uma cadeia composta por 250 resíduos, com 60% de estruturas tipo hélice-α, gerado pelo modelo com distribuição de largura $\delta/\pi = 0.1$. (b) Mapa de Ramachandran para 100 simulações realizadas com os mesmos parâmetros da Figura 3.3 (a)

3.3.1 O raio de giração

A primeira grandeza geométrica relevante na caracterização estrutural das

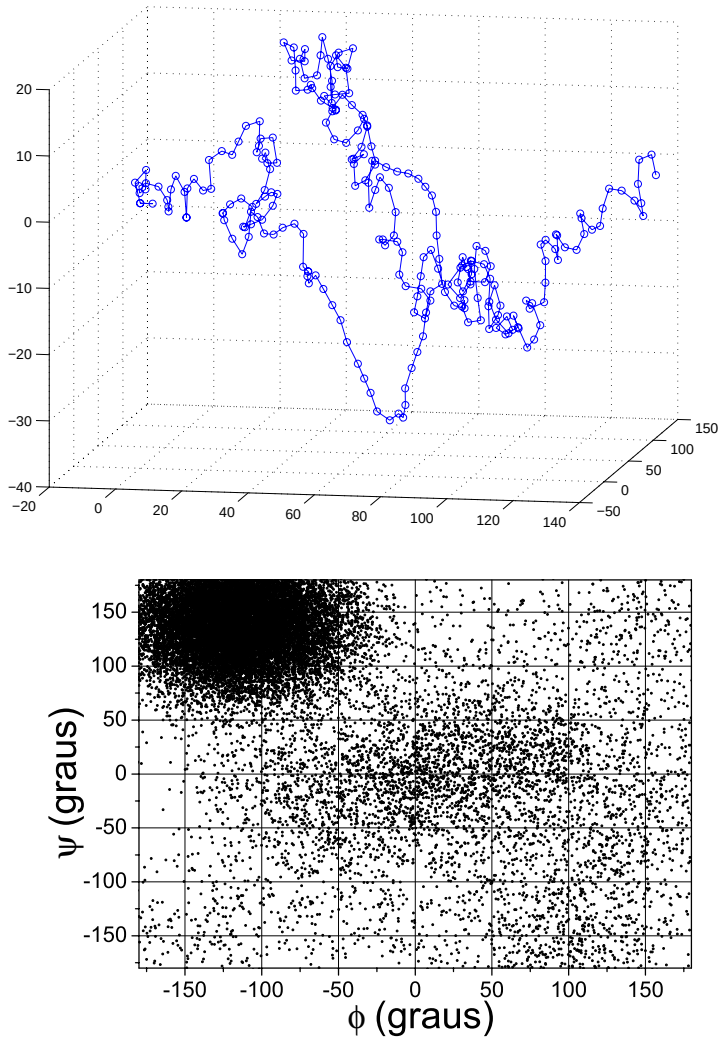

Figura 3.4: (a) Padrão típico de uma cadeia composta por 250 resíduos, com 60% de estruturas tipo folha-β, gerado pelo modelo com distribuição de largura $\delta/\pi = 0.1$. (b) Mapa de Ramachandran para 100 simulações realizadas com os mesmos parâmetros da Figura 3.4 (a)

cadeias protéicas, geradas pelo nosso modelo, é o raio de giração. Assim como definido pela lei de escala descrita na Equação 3.2, o raio de giração determina se o pro-

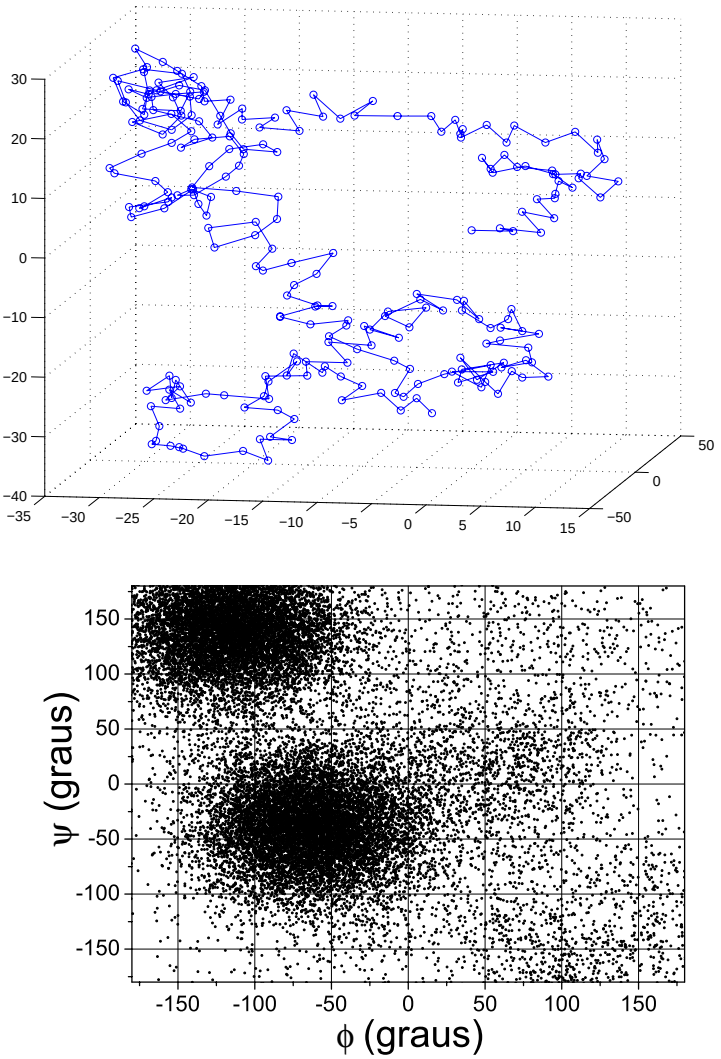

Figura 3.5: (a) Padrão típico de uma cadeia composta por 250 resíduos, com 30% de estruturas tipo hélice-α e 30% de estruturas tipo folha-β, gerado pelo modelo com distribuição de largura $\delta/\pi = 0.1$. (b) Mapa de Ramachandran para 100 simulações realizadas com os mesmos parâmetros da figura 3.5 (a)

cesso de crescimento de nossa cadeia, interpretado como um caminhante aleatório, é sub ou super-difusivo. A Figura 3.6 exibe o comportamento do raio de giração

($< R_g >$), mediado sobre 10^4 estruturas, como função do número de aminoácidos ($125 < N < 450$), para as três diferentes possíveis escolhas de estruturas. Neste estudo, também fixamos a fração $f = 60\%$ e a largura da distribuição angular como sendo $\delta/\pi = 0.1$. Deste gráfico, em escala log-log, podemos obter o expoente ν, o qual fornece uma medida da compactação das cadeias. Seguindo este procedimento encontramos $\nu_\alpha = 0.401 \pm 0.002$ para o caso hélice-α, $\nu_\beta = 0.417 \pm 0.002$ para o caso folha-β e $\nu_{mix} = 0.409 \pm 0.002$ para o caso misturado. Podemos observar, então, que os valores obtidos através dessa metodologia, descrevem de forma razoavelmente satisfatória os resultados experimentais discutidos no gráfico da Figura 3.2. Outro ponto importante é que, independente do tamanho da cadeia gerada (N), as estruturas em hélice-α apresentam-se sempre mais compactas, com menor raio de giração, que aquelas compostas por motivos tipo folha-β, consequência da estrutura tubular, para a primeira e planar para a segunda, como observado por Maritan e colaboradores [91, 92].

Um aspecto importante dessa modelagem é o papel da fração f de estruturas específicas, sejam elas hélice-α, folha-β ou uma mistura de ambas. Inicialmente, utilizamos o valor $f = 60\%$, pois este número reflete a composição média observada nas proteínas. A despeito disso, precisamos investigar em nosso modelo, de que forma o expoente ν se altera com a variação deste parâmetro. Diversas perguntas importantes se colocam: Existiria um valor de f que minimizaria ν, ou seja, que maximizaria a compactação? Este valor é biologicamente compatível? Qual a influência do desvio padrão das distribuições na compactação destas estruturas?

Para tentarmos elucidar estas questões apresentamos na Figura 3.7 um gráfico do comportamento do expoente ν com a fração f, fixando a largura da distribuição angular em $\delta/\pi = 0.1$ e tomando médias sobre 10^4 estruturas. Assim como nos

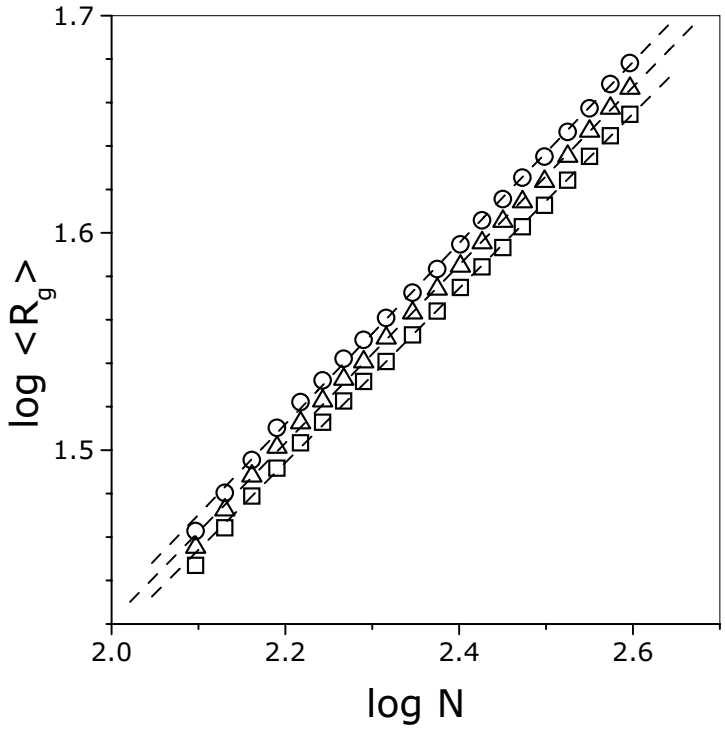

Figura 3.6: Raio de giração médio em função do número de resíduos obtidos por simulação com $f = 0.60$ para estruturas: hélice-α ($\square$), misturadas ($\triangle$) e folhas-β ($\circ$) com expoente de escala 0.401 ± 0.002, 0.409 ± 0.002 e 0.417 ± 0.002, respectivamente. As linhas tracejadas indicam a regressão linear. Em todos os casos as barras de erro são menores que os símbolos e $\delta/\pi = 0.1$.

casos anteriores, investigamos sistemas compostos por estruras secundárias tipo: (a) hélice-α ($\square$); (b) folhas-β ($\circ$) e (c) mistura de hélice-α e folha-β ($\triangle$). A linha tracejada indica o valor experimental $\nu_{exp} \simeq 0.405$. Note que o comportamento monotônico de $\nu(f)$ não fornece uma idéia intuitiva da razão pela qual a maior parte das estruturas correspondem ao valor $f = 0.60$, embora o valor $f = 0.00$ corresponda à máxima compactação. A explicação para este comportamento deve-se à largura δ da distribuição de probabilidades dos ângulos diedrais $P(\phi)$ e $P(\psi)$,

uma vez que ela permite a "visitação" ao espaço de fase das possíveis conformações protéicas.

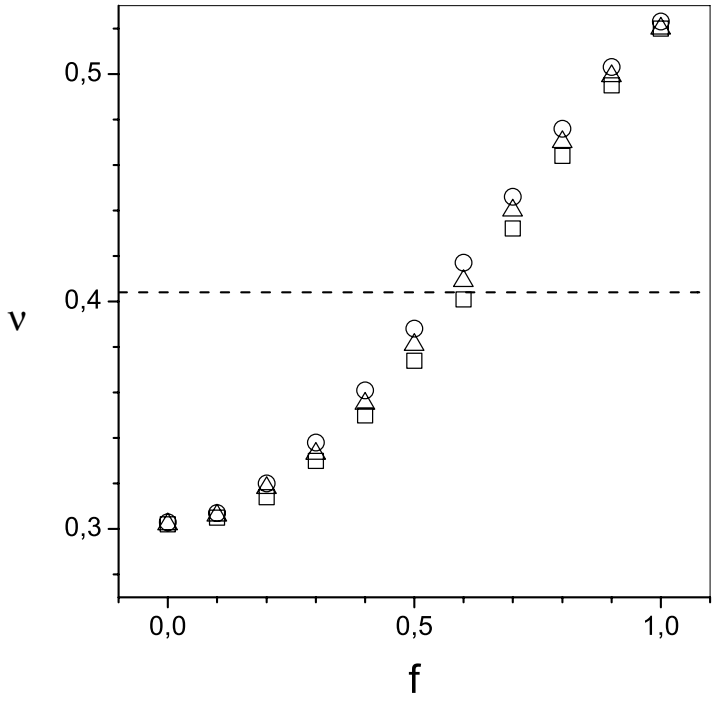

Figura 3.7: Dependência do expoente de escala ν com a porcentagem das estruturas secundárias f para motivos tipo: hélice-αs ($\Box$), folhas-β ($\circ$) e misturadas ($\triangle$). Em todos os casos as barras de erro são menores que os símbolos e $\delta/\pi = 0.1$. A linha tracejada indica o valor experimental $\nu_{exp} \simeq 0.405$.

Na Figura 3.8, exibimos o comportamento do expoente de escala ν como função da largura δ, medida em unidades de π, para diversos valores da fração: (a) $f = 0$ ($\diamond$), (b) $f = 0.40(\bigtriangledown)$, (c) $f = 0.60(\Box)$, (d)$f = 0.80(\circ)$ e (e) $f = 1.00(\triangle)$. Como o comportamento monotônico de ν com f é universal para todas as estruturas estudadas, sejam elas tipo hélice-α, folha-β, ou misturadas, nos concentramos apenas

na análise do caso de cadeias formadas por hélice-αs, uma vez que estas são as mais compactas.

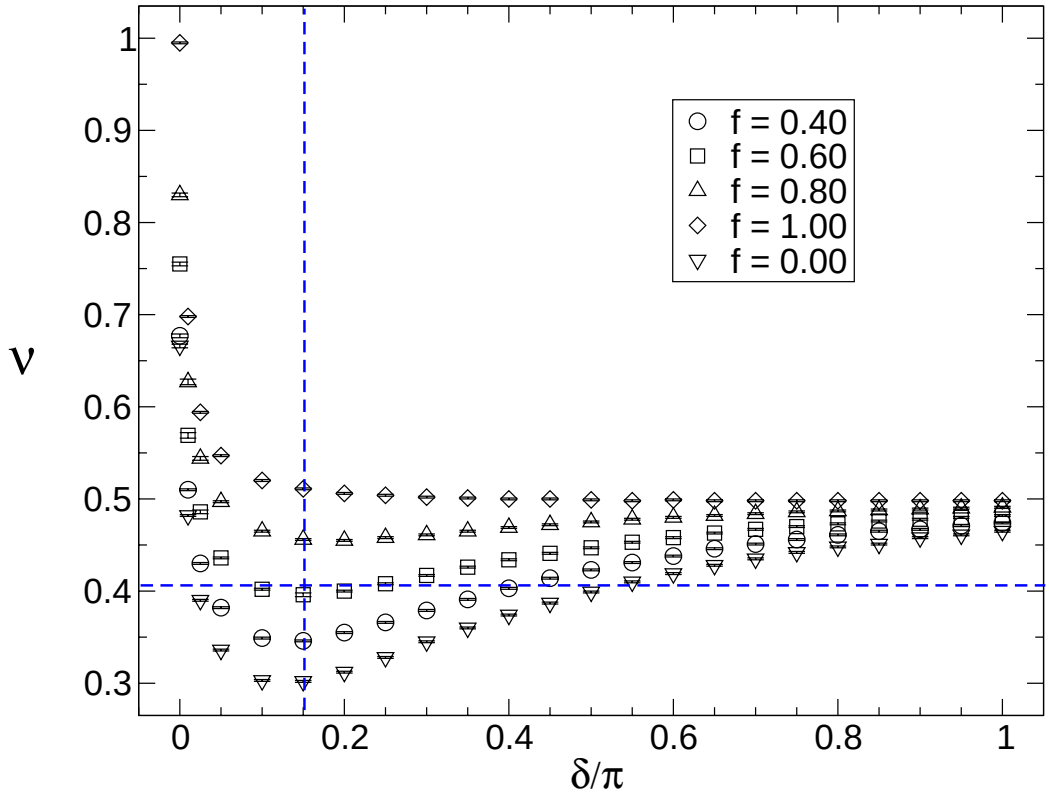

Figura 3.8: Dependência do expoente de escala ν, com a variância δ da distribuição de probabilidade Gaussiana para os ângulos diedrais (em unidades de π), para estruturas em hélice-α. Considerando os valores de $f = 0(\bigtriangledown)$, $f = 0.40(\circ)$, $f = 0.60(\square)$, $f = 0.80(\bigtriangleup)$ and $f = 1.0(\diamond)$. A linha tracejada horizontal indica o valor experimental $\nu_{exp} \simeq 0.405$, enquanto que a linha vertical indica o valor $\delta/\pi = 0.15$, que minimiza ν para qualquer valor de f.

Os resultados expostos no gráfico da Figura 3.8, indicam que para qualquer valor da porcentagem f, de estruturas hélice-α, o valor $\delta_c \approx 0.15\pi$ minimiza o expoente ν, ou seja, maximiza a compactação. Esse resultado mostra também, que no limite $\delta \rightarrow \pi$, o método gera uma estrutura com a caractrística de uma caminhada

aleatória difusiva, com expoente $\nu = 0.5$, para qualquer valor de f, que é esperado, uma vez que neste limite todas as direções são equiprováveis. No limite oposto, quando $\delta \to 0.00$, a estrutura gerada torna-se determinística, com expoente $\nu = 1.0$. Em resumo:

$$\begin{cases} \delta/\pi \to 0 & \nu = 1.0 \\ \delta/\pi \to 1 & \nu = 0.5 \end{cases} \tag{3.8}$$

Embora forneçam uma idéia da robustez e coerência, os resultados obtidos pelo modelo proposto, a despeito do comportamento do expoente ν em função do desvio δ, ainda não justificam de forma direta o percentual observado nos resultados experimentais. É importante notar que nenhum dos ingredientes do modelo indica *a priori* o padrão de compactação das estruturas geradas.

Para investigar a conexão entre a máxima compactação e o parâmetro f, definimos um parâmetro geométrico, que mede quão representativo é o raio de giração em termos da máxima dimensão do sistema. Denominamos essa grandeza de "parâmetro de compactação":

$$\gamma \equiv \frac{R_g}{D_{max}}, \tag{3.9}$$

onde R_g é o raio de giração da estrutura e D_{max} é a distância entre os resíduos mais distantes na estrutura. Na Figura 3.9 exibimos, esquematicamente, o exemplo do raio de giração e a máxima distância para uma estrutura bidimensional arbitrária.

Fixando o tamanho das estruturas geradas em $N = 250$ resíduos e variando a fração $0 < f < 1$, geramos cadeias compostas por estruturas secundárias formadas

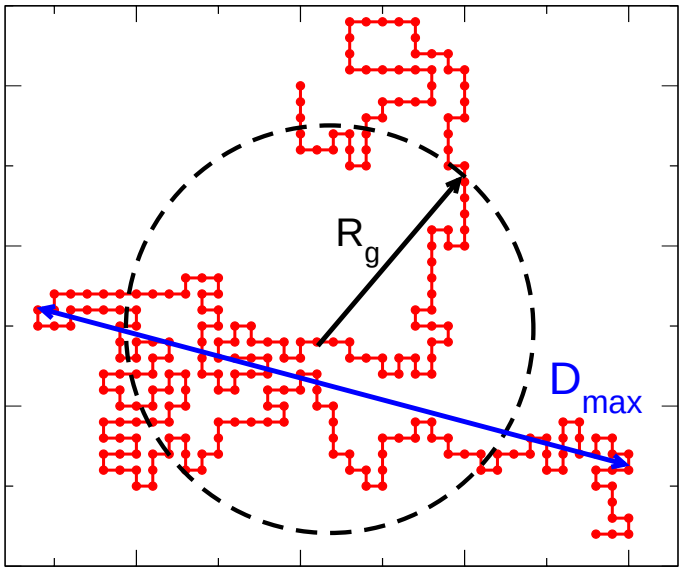

Figura 3.9: Grandezas envolvidas na determinação do "parâmetro de compactação" γ para uma estrutura bidimensional arbitrária. Raio de giração R_g (preto) e distância máxima D_{max} (azul).

por motivos tipo hélice-α, folha-β e uma mistura equitativa desses dois motivos. Na Figura 3.10 exibimos o comportamento de γ, mediado sobre 10^4 conformações, para cada valor de f (nestas simulações utilizamos $\delta/\pi = 0.15$).

Inicialmente é importante registrar que mesmo sendo percentualmente pequena (da ordem de $2,22\%$), é possível discernir dentro da barra de erro dos resultados uma variação do parâmetro γ com a fração f. Este comportamento exibe claramente um máximo em torno de $f = 0.60$, sendo similar para qualquer tipo de estrutura secundária característica utilizada. Note ainda que, assim como já indicado pelo expoente de escala ν, estruturas tipo hélice-αs possuem o maior parâmetro de compactação $\gamma^{\alpha}_{max} = 0.3215 > \gamma^{mix}_{max} = 0.3210 > \gamma^{\beta}_{max} = 0.3205$. Assim, de acordo com o modelo proposto, o valor $f = 0.60$ é aquele que determina estruturas mais compactas, no sentido de que a proporção entre as maiores dimensões do sistema são

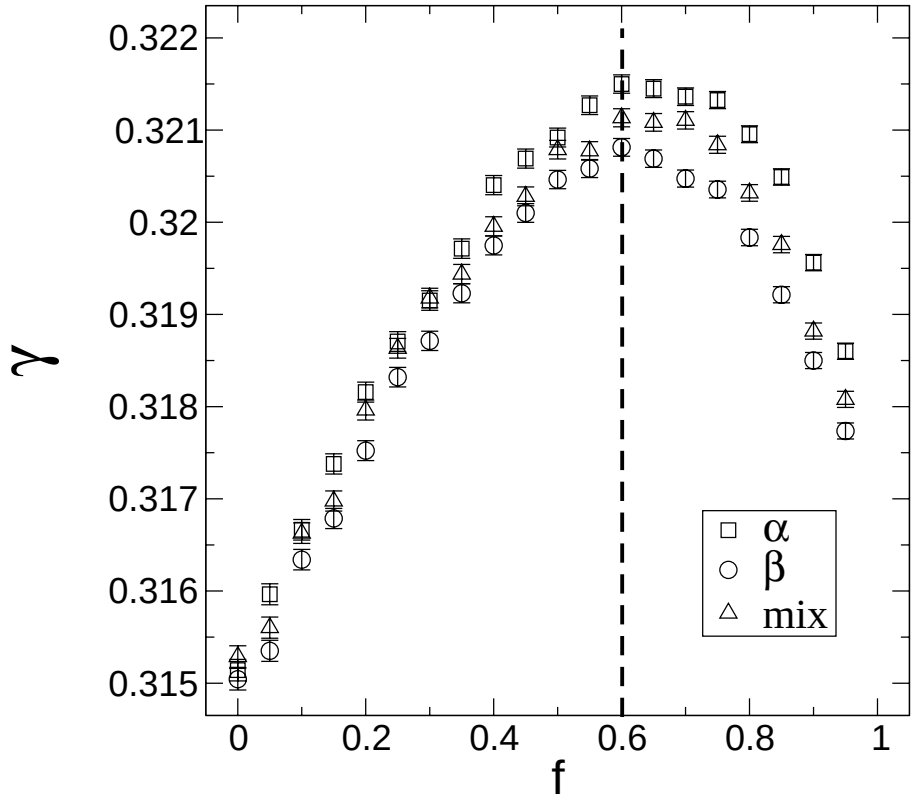

Figura 3.10: Dependência do parâmetro γ (mediado sobre 10^4 amostras) com a porcentagem de estruturas secundárias f para motivos tipo: hélice-αs ($\Box$), folhas-β ($\circ$) e misturadas ($\triangle$). A linha tracejada indica o valor máximo γ_{max}, em todos os casos, para $f = 0.60$. A cadeia inteira possui $N = 250$ resíduos e largura $\delta/\pi = 0.15$.

as mais próximas possíveis do raio de giração do mesmo. Para corroborar a robustez deste resultado apresentamos na Figura 3.11 um histograma do parâmetro γ, calculado segundo a definição (3.9), para 1356 diferentes cadeias globulares, extraídas do PDB, com número de resíduos variando entre $125 < N < 450$. Para este conjunto de dados o tratamento estatístico revela que: $\gamma_{exp,min} = 0.249$, $\gamma_{exp,max} = 0.373$ e que o valor médio dessa distribuição é $\gamma_{exp} = 0.32 \pm 0.02$; valores muito próximos àqueles obtidos através do modelo proposto. É importante notar que o valor de γ decresce

ligeiramente com o comprimento N da cadeia; de modo que os dois picos presentes na distribuição apresentada na Figura 3.11 justificam-se pela presença de estruturas com tamanhos N específicos, como pode ser observado na distribuição do tamanho das cadeias presentes no banco de dados utilizado, exposta na Figura 3.12. Este comportamento também é qualitativamente reproduzido para o modelo proposto, onde os dois picos mais pronunciados correspondem aos valores $N = 162$ e $N = 372$, como pode ser verificado no histograma para o parâmetro γ exibido na Figura 3.13. Neste caso, calculamos γ para as 1356 diferentes estruturas globulares utilizadas no histograma da Figura 3.11, através do modelo proposto com $\delta/\pi = 0.15$ e com $f = 60\%$ de estruturas tipo hélice-α.

3.3.2 O comprimento de contorno

Outra análise comumente utilizada na caracterização estrutural de cadeias protéicas é a do comportamento de "comprimento de contorno", l_{ij}, definido como a distância, medida ao longo da cadeia, que separa dois resíduos [93]; em função da distância direta entre estes os mesmos r_{ij}, como nos mostra esquematicamente a Figura 3.14. De forma prática são computados valores médios destas duas grandezas, mediados por todos os pares de carbonos C_α dentro de uma dada estrutura.

O comportamento do comprimento de contorno $< l_c >$ com relação a distância Euclideana r nos fornece uma idéia da estrutura topológica do objeto estudado, uma vez que para um sistema com a topologia de uma linha, por exemplo teremos um comportamento linear entre estas grandezas. Uma lei de escala entre as mesmas,

$$< l_c > \sim r^{\eta}, \tag{3.10}$$

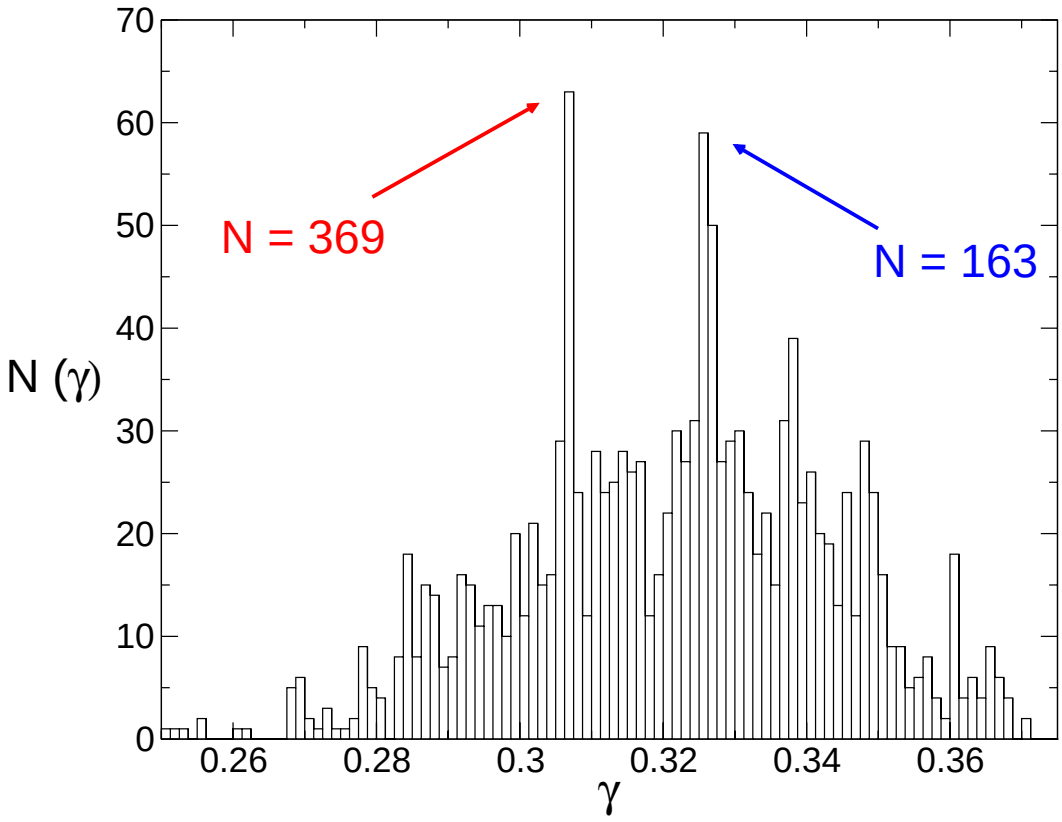

Figura 3.11: Histograma do parâmetro γ calculado para 1356 diferentes estruturas globulares com $125 < N < 450$ resíduos, extraídas do PDB. Valor médio da distribuição $\gamma_{exp} = 0.32 \pm 0.02$. Os dois picos mais pronunciados correspondem aos valores $N = 163$ e $N = 369$.

indicaria em média a distância efetiva entre dois elementos da cadeia. Como já observado na seção anterior, a grandeza γ, que mede uma possível compactação do sistema não sofre alteração apreciável com o tamanho da cadeia, desta forma a fim de estudar a lei de potência proposta na Equação 3.10 vamos considerar um número fixo de carbonos C_α ($N = 300$ aminoácidos) e variar apenas a fração f quanto a largura δ de nossa distribuição angular.

Na Figura 3.15, exibimos o comportamento típico do comprimento de con-

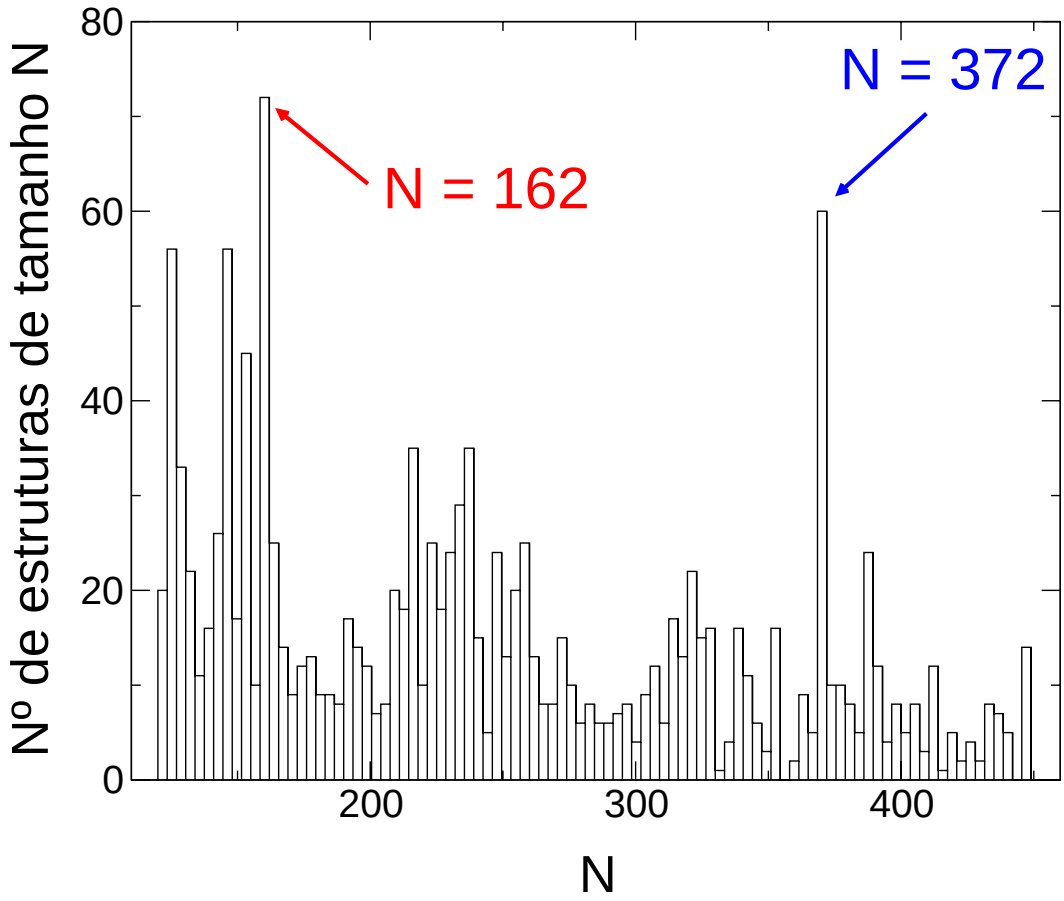

Figura 3.12: Histograma do tamanho N das 1356 diferentes estruturas globulares utilizadas ao longo deste Capítulo. Os dois picos mais pronunciados correspondem aos valores $N = 162$ e $N = 372$.

torno $< l_c >$, como função da distância direta $< r >$, para uma média de 10^4 amostras, compostas por motivos com estrutura secundária tipo hélice-α, para diferentes valores da largura da distribuição δ/π e para $f = 0.00$. Na Figura 3.16, realizamos o mesmo estudo utilizando agora $f = 0.60$. Como podemos perceber deste resultados, o comportamento tipo lei de potência sugerida na Equação 3.10 é obtido, com uma clara mudança do expoente η, para distintos valores de δ/π. Por outro lado observamos também que o parâmetro f, não controla o comportamento

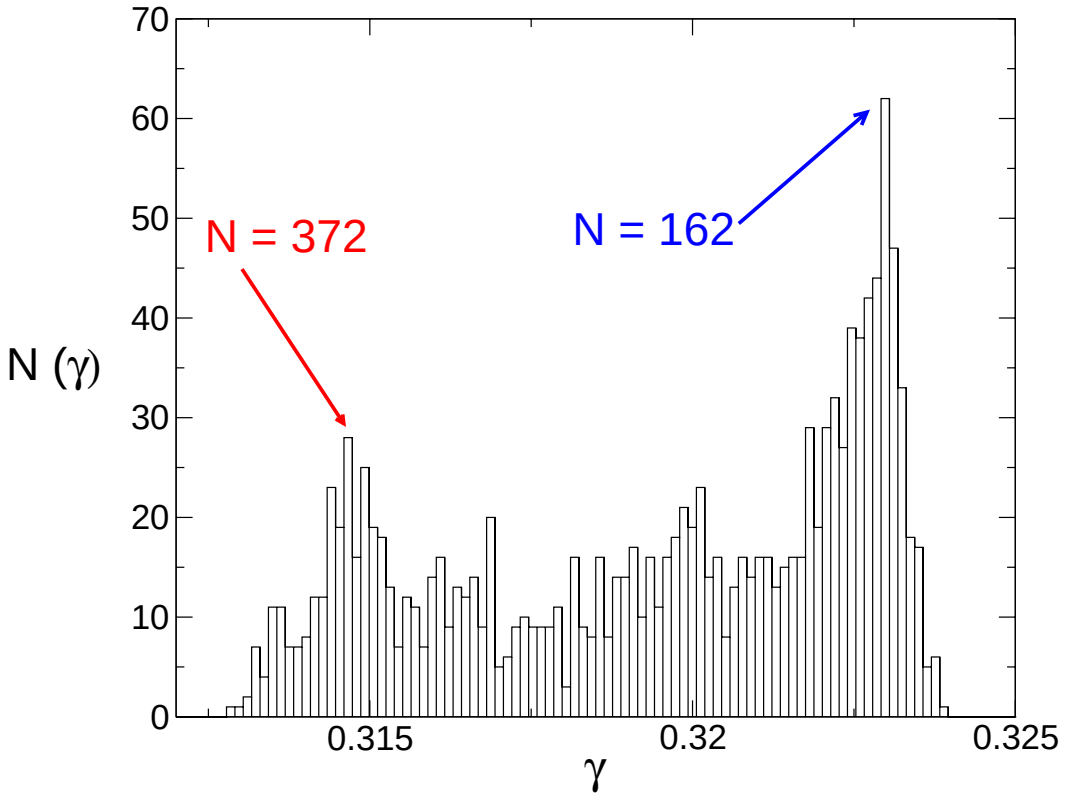

Figura 3.13: Histograma do parâmetro γ calculado para as 1356 diferentes estruturas globulares, utilizadas no histograma da Figura 3.11, através do modelo proposto com $\delta/\pi = 0.15$ e com $f = 60\%$ de estruturas tipo hélice-α. Valor médio da distribuição $\gamma_{f=0.60} = 0.32 \pm 0.02$. Os dois picos mais pronunciados correspondem aos valores $N = 162$ e $N = 372$. O valor de γ das estruturas simuladas é determinado por médias para 10^4 estruturas similares aquelas reais.

de η, evidenciando que diferentemente do expoente ν, para o caso do raio de giração, o comportamento tipo lei de potência depende intrínsecamente aleatoriedade do sistema relacionado com δ/π.

Para confirmarmos as afirmações acima, realizamos um grande número de simulações, variando δ/π e f. Estes resultados encontram-se na Figura 3.17, onde apresentamos o comportamento de η como função de δ/π para: $f = 0.00\,(\circ)$, $f =$

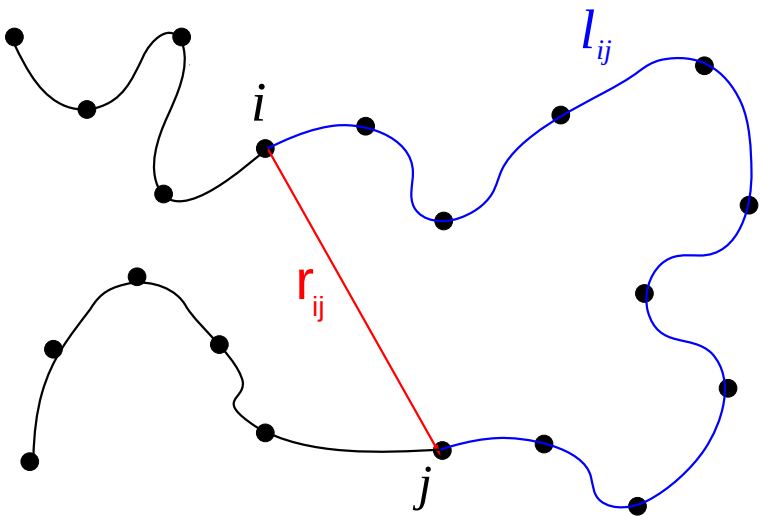

Figura 3.14: Figura esquemática exemplificando o cálculo da distância direta r (vermelho) e do comprimento de contorno l_{ij}, entre dois elementos (azul) de uma estrutura bidimensional arbitrária.

$0.60\,(\square)$ e $f = 1.00\,(\triangle)$. Para cada ponto realizamos 10^4 amostras e fixamos o número de resíduos em $N = 300$, o comportamento assintótico para $\delta/\pi \to 0$ indica que $(\eta \to 1)$, como comentado no início desta seção, uma vez que apenas um ângulo é sorteado na construção do caminhante. No outro limite quando $\delta/\pi \to 1$ todos os ângulos são possíveis e retomamos o caso de um caminahnte aleatório usual $(\eta \to 2)$.

3.3.3 O número de coordenação e a energia de contato.

Embora o raio de giração seja um parâmetro macroscópico largamente utilizado na caracterização do enovelamento protéico, ele é uma grandeza geométrica, que não leva em consideração a sequência de aminoácidos que compõe uma proteína específica. Este é um ponto relevante para o estudo de conformações de uma cadeia proteíca, uma vez que resíduos espacialmente próximos, mas não ligados diretamente ao longo da cadeia, interagem fisicamente, estabelecendo ligações efetivas e

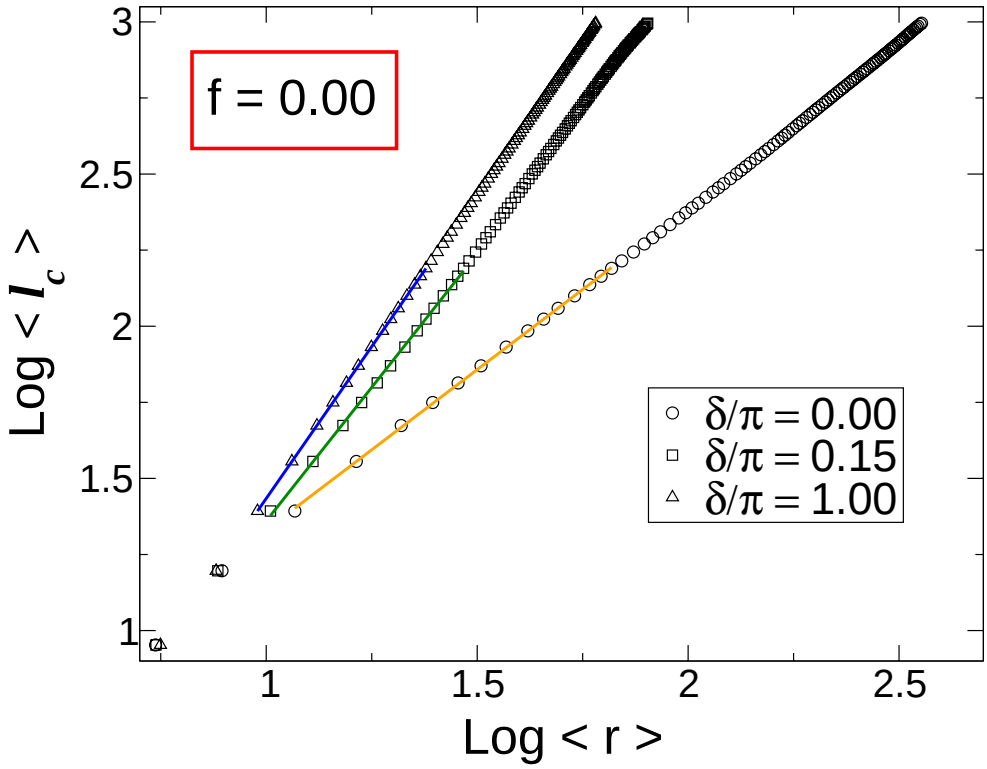

Figura 3.15: Comportamento do comprimento de contorno $< l_c >$, como função da distância direta r, mediado sobre 10^4 amostras, e com uma variação da largura da distribuição angular para três valores $\delta/\pi = 0.00(\circ)$, $\delta/\pi = 0.15$ $(\square)$ e $\delta/\pi = 0.15$ $(\triangle)$. Neste caso fixamos a fração de estruras típicas $f = 0.00$.

contribuindo decisivamente para a estabilização do estado nativo. Como discutido por Levitt e colaboradores [87], a maior parte das funções de energia utilizadas para prever conformações protéicas depende de estimativas sobre o número e o tipo de contato entre os resíduos. Ademais, a descoberta de que um pequeno número de contatos, formados nos primeiros estágios do processo de enovelamento, são decisivos para a estabilização da estrutura final da proteína, indica a necessidade de um estudo mais aprofundado do número de contatos e da energia entre estes contatos, para o modelo. Para determinarmos o número de contatos em uma dada estrutura

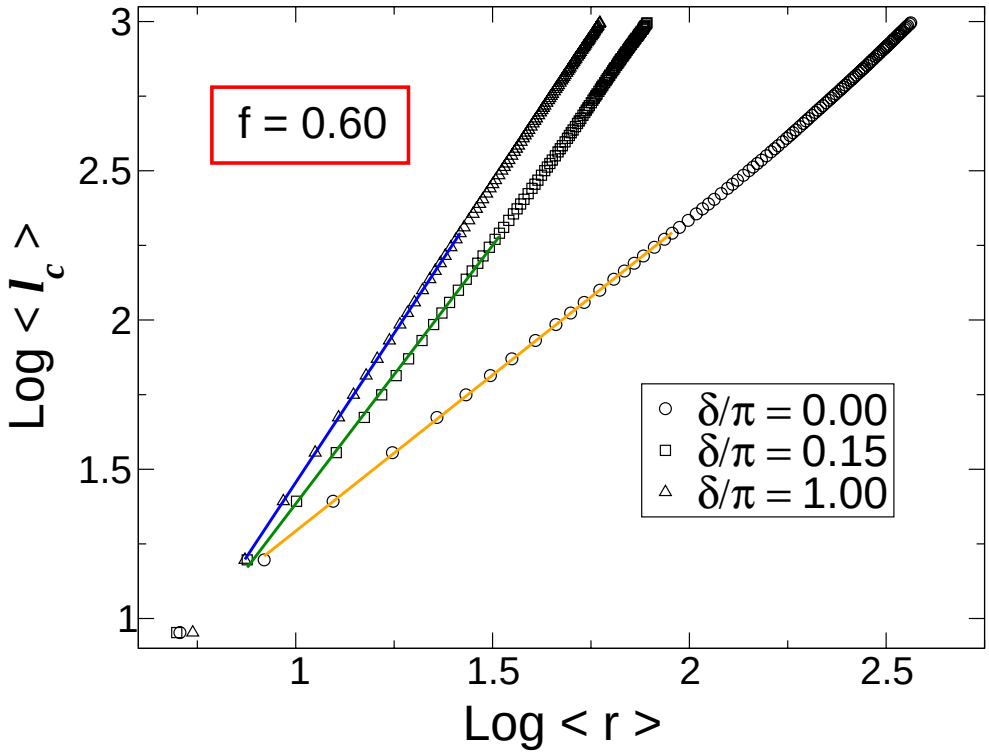

Figura 3.16: Comportamento do comprimento de contorno $< l_c >$, como função da distância direta r, com os mesmos parâmetros da figura 3.15. Aqui fixamos a fração de estruturas típicas $f = 0.60$.

procedemos da seguinte forma:

1. Construímos uma estrutura composta por N resíduos (designados por seus carbonos $C_{i\alpha}\, i = \{1, ..., N\}$) determinando as coordenadas cartesianas de cada um deles.

2. Tomando uma esfera de raio r_c, centrada num dado resíduo k, computamos o número de resíduos, não sequenciais dentro deste raio (para ilustração ver a Figura 3.18).

3. Repetimos o procedimento para cada resíduo. Excluindo a possibilidade de

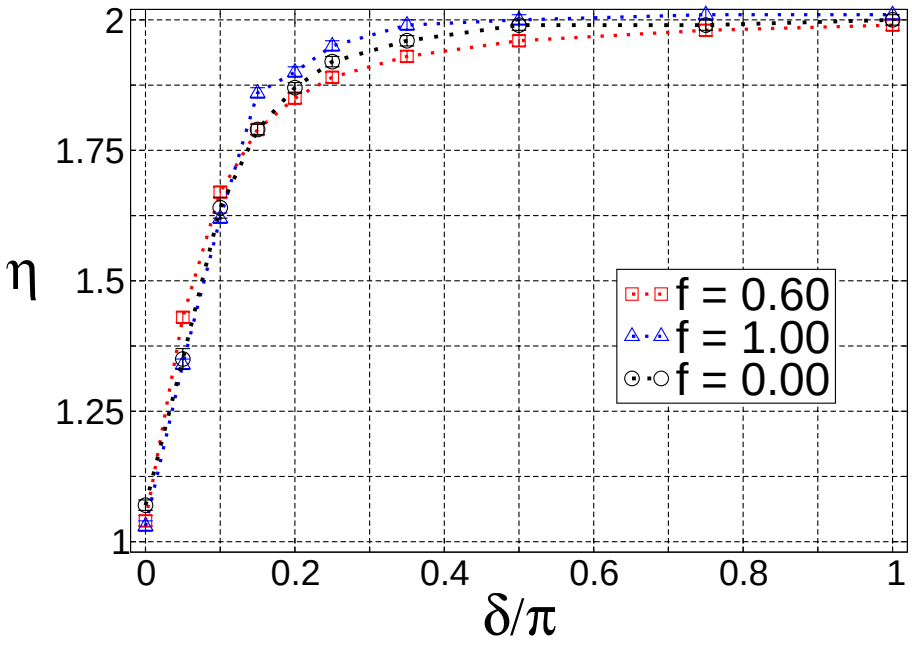

Figura 3.17: Comportamento de η como função de δ/π para $f = 0.00\,(\circ)$, $f = 0.60\,(\square)$ e $f = 1.00\,(\triangle)$. Para cada ponto realizamos 10^4 amostras e fixamos o número de resíduos em $N = 300$.

computarmos o mesmo par duas vezes.

$$n_c = \sum_{k=1}^{N} \sum_{\substack{j=1 \\ (j\neq k, j\neq k+1, j\neq k-1)}}^{N} \Delta(\mathbf{r}_k - \mathbf{r}_j), \tag{3.11}$$

onde a função $\Delta(\mathbf{r}_k - \mathbf{r}_j)$ é definida por:

$$\Delta(\mathbf{r}_k - \mathbf{r}_j) \equiv \begin{cases} 0 & \text{se } |(\mathbf{r}_k - \mathbf{r}_j)| > r_c \\ 1 & \text{se } |(\mathbf{r}_k - \mathbf{r}_j)| \leq r_c, \end{cases} \tag{3.12}$$

$\mathbf{r}_k$ e $\mathbf{r}_j$ são, respectivamente, as posições do k-ésimo e do j-ésimo carbono alfa da estrutura, e r_c é o raio de contato característico para motivos tipo hélice-α e folha-β $r_c = 7$Å. Conjugada à definição do número de contatos também utilizamos, como proposto por Tang e colaboradores [90], o número médio de coordenação z_c para uma estrutura, definido por:

$$< z_c > \equiv \frac{n_c}{N}, \tag{3.13}$$

como forma de caracterizar o número médio de contatos de um dado resíduo.

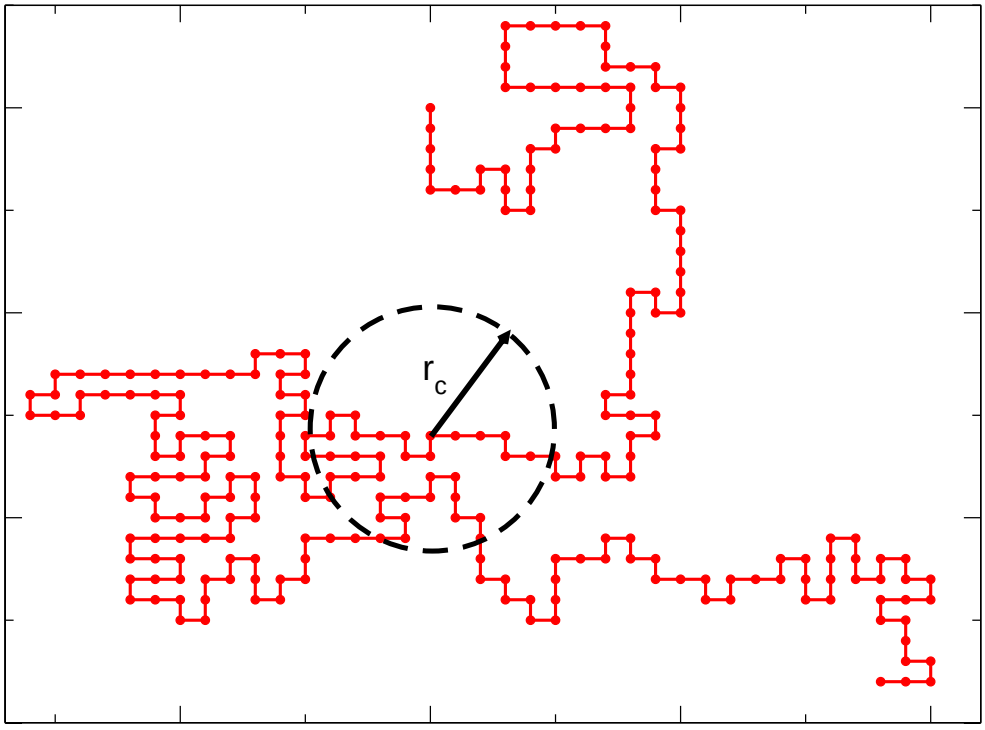

Figura 3.18: Exemplo do cálculo do número de contatos para uma estrutura bidimensional arbitrária. Nesta figura, o elemento indicado possui 28 contatos.

Estabelecida a metodologia empregada, podemos passar à análise dos resulta-

dos obtidos para o número de contatos. Inicialmente verificamos o comportamento do número de contatos n_c como função do número de resíduos na estrutura N. Neste estudo, inicialmente fixamos o valor de $\delta/\pi = 0.15$ (valor que maximiza o "parâmetro de compactação" γ) e variamos a fração f de estruturas secundárias. Na Figura 3.19 exibimos o comportamento do número de contatos n_c como função de N, enquanto que na Figura 3.20 o número de coordenação médio z_c, nesses casos utilizamos como estruturas secundárias apenas motivos tipo hélice-αs.

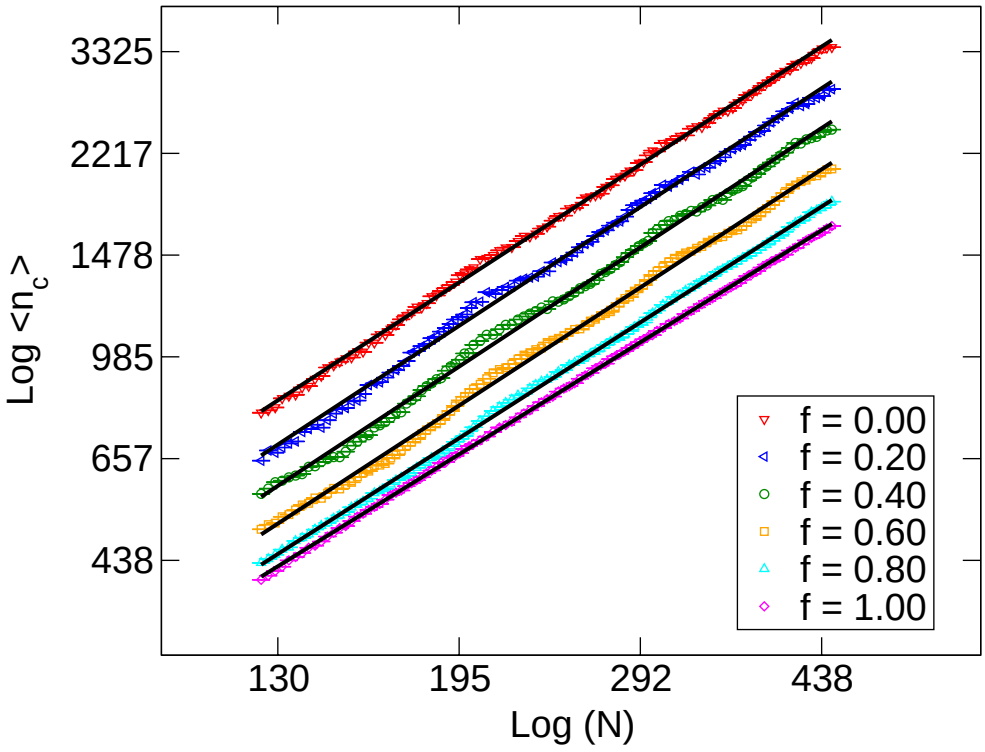

Figura 3.19: Comportamento do número de contatos $< n_c >$, como função do comprimento da cadeia N, para diversos valores da fração f. Em todas as simulações utilizamos como motivos apenas estruturas tipo hélice-α e mediamos sobre 10^4 amostras. As retas em preto são ajustes seguindo o comportamento de escala proposto na Equação 3.14.

Como observado na literatura [90] o comportamento do número de contatos

n_c é praticamente linear, como o número de resíduos N, de forma que uma relação de escala do tipo:

$$< n_c > \sim N^{\chi}, \tag{3.14}$$

fornece um expoente χ muito próximo a $\chi = 1.00$, como podemos observar na Tabela 3.2.

f	χ
0.00	1.16 ± 0.01
0.20	1.17 ± 0.01
0.40	1.17 ± 0.01
0.60	1.16 ± 0.01
0.80	1.14 ± 0.01
1.00	1.10 ± 0.01

Tabela 3.2: Valores dos expoentes χ e respectivos desvios obtidos através da relação de escala definida na Equação 3.14, como função da fração f

O comportamento linear do número de contatos, é melhor observado para sistemas com maior número de resíduos presentes na cadeia protéica, de forma que o número médio de coordenação $< z_c >$ apresenta uma saturação para grandes valores de N, como evidenciado na Figura 3.20.

Assim como no estudo do "parâmetro de compactação" γ, procuramos comparar os resultados obtidos para o número de coordenação, através de nossas simulações com dados experimentais. Para tanto selecionamos 1356 estruturas do PDB, como critério procuramos cadeias globulares, possuindo um número $125 < N < 450$ de aminoácidos. Destas, isolamos apenas seus carbonos C_α. De posse do tipo dos aminoácidos e das coordenadas destes carbonos, pode-se calcular o número

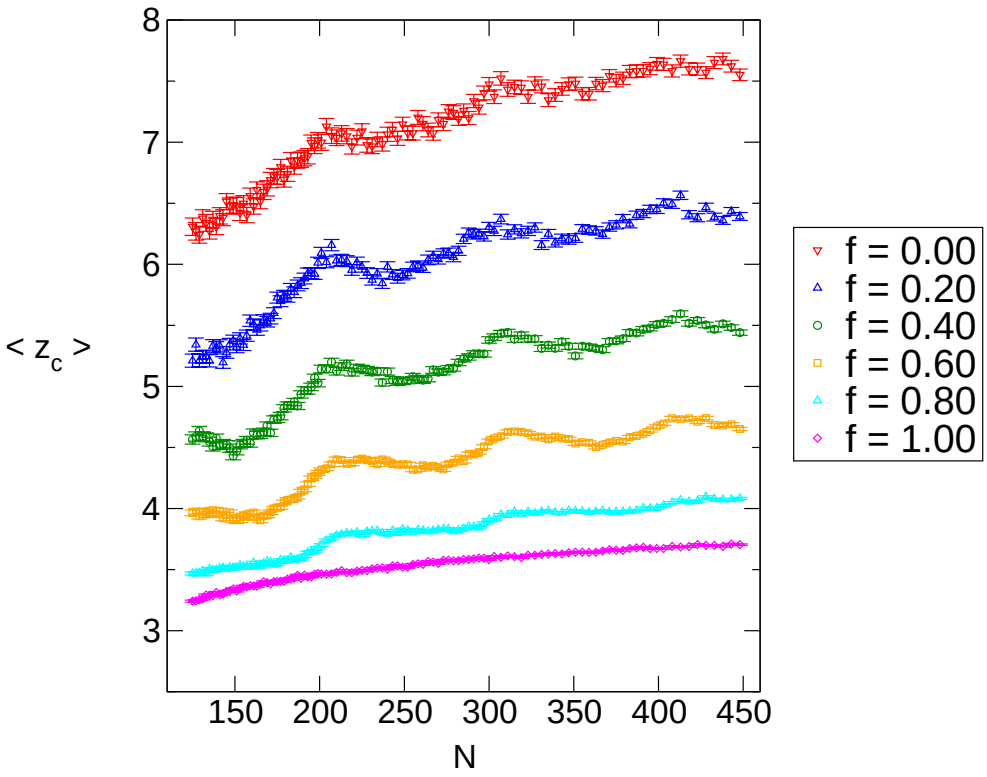

Figura 3.20: Comportamento do número médio de coordenação $< z_c >$ como função do comprimeto da cadeia N, para os mesmos parâmetros utilizados na Figura 3.19. Observe o comportamento de saturação para grandes valores de N.

de coordenação para estas estruturas, por meio da metodologia discutida anteriormente, e comparar com aquelas geradas pelo modelo proposto. Na Figura 3.21 confrontamos diretamente o número de coordenação calculado para todos os sistemas do banco de dados (em preto), e seus equivalentes simulados, cada um mediado sobre 10^4 estruturas, para distintos valores de f = (vermelho), $f = 0.20$ (azul), $f = 0.40$ (verde), $f = 0.60$ (laranja), $f = 0.80$ (ciano) e $f = 1.00$ (lilás). Tanto nas simulações, quanto no caso das estruturas reais utilizamos raio de contato $r_c = 7$Å. No caso dos resultados do modelo, fixamos $\delta/\pi = 0.15$ e utilizamos uma mistura de motivos α-hélcie e folha-β.

Os resultados, descritos na Tabela 3.3, apontam que estatisticamente, a máxima sobreposição entre a distribuição experimental e aquelas simuladas ocorreria para $0.60 < f < 0.80$, em concordância com a predição obtida através do "parâmetro de compactação" γ, isto é a fração $f \cong 0.60$ é aquela que melhor caracteriza as estruturas reais. Em particular $n_c^{PDB} = 4.0 \pm 0.2$ e $n_c^{f=0.60} = 4,3 \pm 0.2$.

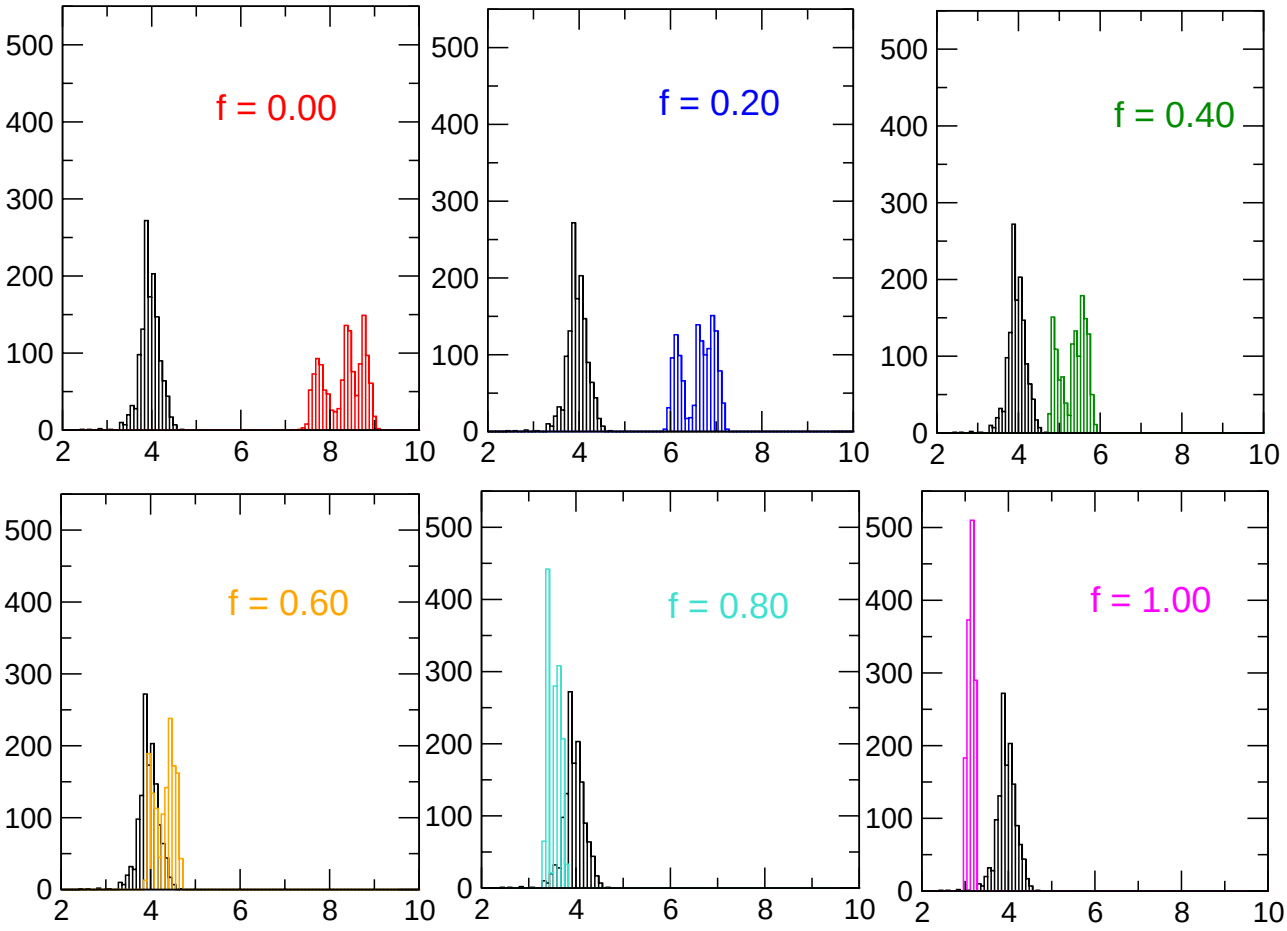

Figura 3.21: Distribuições dos valores do número de coordenação z_c obtidas pelo modelo para diversos valores da fração $f = 0$ (vermelho), $f = 0.20$ (azul), $f = 0.40$ (verde), $f = 0.60$ (laranja), $f = 0.80$ (ciano), $f = 1.00$ (lilás) e para 1356 diferentes estruturas extraídas do PDB (preto). Em todas as simulações utilizamos 10^4 amostras, largura $\delta/\pi = 0.15$ e raio de contato $r_c = 7$Å.

f	$< z_c >$
0.00	8.3 ± 0.4
0.20	6.6 ± 0.4
0.40	5.3 ± 0.3
0.60	4.3 ± 0.2
0.80	3.5 ± 0.1
1.00	3.1 ± 0.1
PDB	4.0 ± 0.2

Tabela 3.3: Valores dos números de coordenação e respectivos desvios para as distribuições da Figura 3.21, como função da fração f, e para 1356 estruturas do PDB

Até o momento discutimos aspectos geométricos das cadeias geradas através de um modelo de caminhante aleatório minimalista. Vimos que a relação de escala entre o raio de giração dessas estruturas, está muito próxima àquela obtida experimentalmente. Observamos que o número de médio de coordenação também consegue capturar as características básicas de sistema reais e que a fração de estruturas secundárias $f \cong 0.60$, indica a situação de máxima compactação para todas as estruturas. Contudo o modelo não se fundamenta numa dinâmica de minimização de energia, ou seja, não foi expolrado o espaço das possíveis configurações em busca de uma configuração mais estável. As configurações são obtidas por meio de potenciais implicitamente descritos apenas pelos ângulos diedrais Φ e Ψ. Neste ponto, uma pergunta pertinente pode ser colocada: as conformações geradas por este modelo seriam capazes de incorporar de fato as interações entre os aminoácidos que formam uma cadeia protéica? Para responder a esta pergunta precisamos saber se estruturas geradas por este modelo encontram similares de energia em cadeias protéicas reais.

Nesta abordagem utilizamos a seguinte metodologia: partindo das mesmas cadeias utilizadas para o estudo do número de coordenação, construímos, para cada

cadeia do banco de dados (PDB), 10^4 estruturas através de nosso modelo decorando-as com sua sequência primária de aminoácidos. Para o processo de "decoração" de cada configuração, utilizamos os valores contidos na Tabela V proposta por Miyazawa e Jernigan [59], classicamente utilizada nesse tipo de abordagem [73]. Para determinarmos a energia de contato das estruturas, utilizamos uma definição que leva em conta apenas os aminoácidos que estão em contato, como definido na Equação 3.12, desta forma:

$$E = \frac{1}{N}\sum_{k=1}^{N} \sum_{\substack{j=1 \\ (j\neq k, j\neq k+1, j\neq k-1)}}^{N} \varepsilon_{\lambda_k\lambda_j}\Delta(\mathbf{r}_k - \mathbf{r}_j), \tag{3.15}$$

onde a função $\Delta(\mathbf{r}_k - \mathbf{r}_j)$ é definida como na Equação 3.12, $r_c = 7.0$Å, λ_k é o tipo de aminoácido que se encontra na posição $\mathbf{r}_k$, λ_j é o tipo de aminoácido que se encontra na posição $\mathbf{r}_j$, $\varepsilon_{\lambda_k\lambda_j}$ é a energia de interação entre estes aminoácidos, medida em unidades de RT, como estabelecido em [59] e N é o tamanho da cadeia.

Isto posto, podemos analisar a distribuição de energia para as diversas estruturas geradas com a variação da fração f, e comparar estes resultados com a energia computada a partir das cadeias originais contidas no banco de dados (PDB). Para estas simulações utilizamos $\delta/\pi = 0.15$ e motivos compostos por uma mistura equitativa de estruturas secundárias tipo hélice-α e folhas-β. Na Figura 3.22 exibimos o comportamento das distribuições de energia encontradas seguindo a metodologia acima descrita em cada um dos gráficos comparamos diretamente a distribuição obtida através das cadeias reais (preto) com aquelas obtidas por simulação para diversos valores de $f = 0$ (vermelho), $f = 0.20$ (azul), $f = 0.40$ (verde), $f = 0.60$ (laranja), $f = 0.80$ (ciano) e $f = 1.00$ (lilás). A comparação entre gráficos, mais uma vez indica que o valor $f = 0.60$ é aquele que melhor descreve os resultados

experimentais. Uma análise estatística, comparando as distribuições experimental (preto-PDB) e simuladas para $f = 0.60$ (em laranja), revela que o valor médio $E_{PDB} = -11.4 \pm 0.81\,(u.a.)$ enquanto $E_{f=0.60} = -9.84 \pm 0.71\,(u.a.)$. A Tabela 3.22 exibe a dependência da energia $E(u.a.)$ como função da fração f de estruturas em hélice-α eou folha-β. É importante notar que neste estudo utilizamos cadeias com número de resíduos distintos, o que mais uma vez corrobora a robustez do modelo.

f	$E\,(u.a.)$
0.00	-22.0 ± 2.0
0.20	-17,7 ± 1.6
0.40	-14.3 ± 1.3
0.60	-11.6 ± 1.0
0.80	-9,5 ± 0.7
1.00	-8,4 ± 0.6
PDB	-11,1 ± 0.8

Tabela 3.4: Valores das energias médias e desvios para as distribuições da Figura 3.22, como função da fração f, e para as 1356 estruturas do PDB

Sumarizando, os resultados expostos nesse capítulo, revelam que um modelo minimalista, para a construção de cadeias protéicas, baseado num caminhante alatório, cujos ângulos de deflexão seguem uma distribuição de probabilidades Gaussiana, fornece uma descrição qualitativa muito próxima àquela dos dados experimentais. Obtivemos particularmente boas concordâncias para: o expoente de escala ν, o número de contatos n_c, o número de coordenação z_c e para a energia de contato E. Resaltamos, ainda, que o fator decisivo na construção destas estruturas, está relacionado com a estocasticidade do modelo, aqui representada pela largura da distribuição de probabilidades angular δ, ou seja, segundo o modelo proposto as flutuações exerceriam um papel fundamental na estabilidade destas conformações.

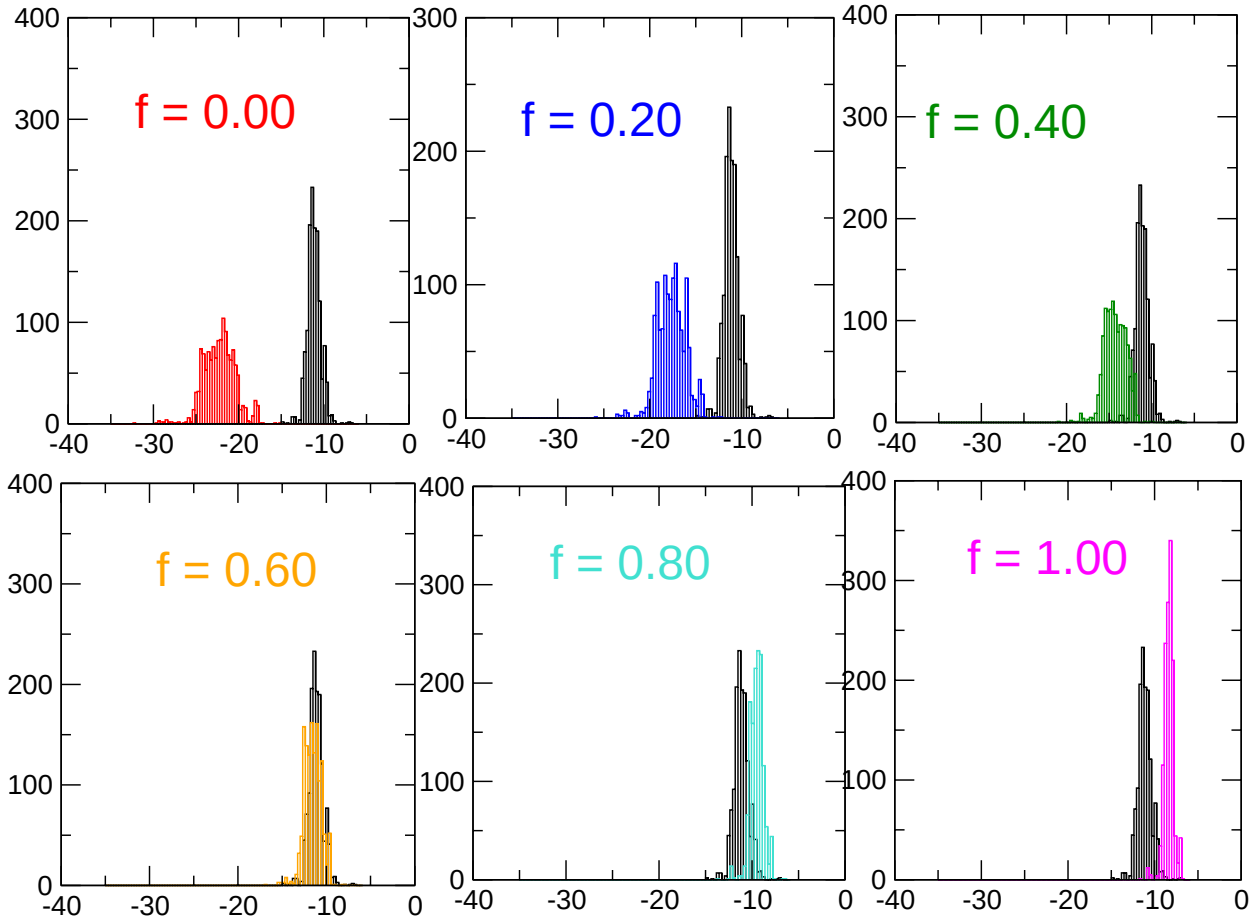

Figura 3.22: Histogramas das distribuições de energia (em u.a.) obtidas pelo modelo para diversos valores da fração $f = 0$ (vermelho), $f = 0.20$ (azul), $f = 0.40$ (verde), $f = 0.60$ (laranja), $f = 0.80$ (ciano), $f = 1.00$ (lilás) e para 1356 diferentes estruturas contidas no PDB (preto). Em todas as simulações utilizamos 10^4 amostras, $\delta/\pi = 0.15$ e raio de contato $r_c = 7$Å.

Capítulo 4

Aspectos multifractais de séries temporais da energia potencial de polipeptídeos

"O mundo todo marcado à ferro, fogo e desprezo. A vida é o fio do tempo, a morte o fim do novelo. O olhar que assusta anda morto. O olhar que avisa anda aceso. Mas quando eu chego eu me perco. Nas tramas do teu segredo."

Dori Caymmi e Paulo César Pinheiro - Desenredo

4.1 A energia potencial de proteínas

No Capítulo anterior, estudamos características estruturais de proteínas com um modelo que se propunha a descrever o estado enovelado de cadeias protéicas, sem nos preocuparmos em expor os mecanismos dinâmicos que levariam o sistema de seu estado desenovelado, para seu estado nativo. É sabido que tais mecanismos

envolvem um profundo entendimento das interações entre os aminoácidos, presentes em sua estrutura primária, e as moléculas que compõem o solvente, onde o sistema encontra-se embebido. Por outro lado, o estado enovelado representa para o sistema uma condição de estabilidade termodinâmica, onde se presume, que as atividades biológicas da protéina se desenvolvam.

O principal obstáculo no estudo das conformações intermediárias que conduzem o sistema ao seu estado enovelado (nativo), deve-se ao significativo número de graus de liberdade envolvidos no processo. Este fato induz o aparecimento de inúmeros estados de mínimas energias, próximos ao mínimo global do sistema. Isso implica que as hiper-superfícies de energia potencial são N-dimensionais e não convexas. Desta forma as investigações destas hiper-superfícies só podem ser feitas numericamente. Entre os métodos utilizados nas simulações computacionais destacamos: o método de Monte Carlo (MC); a Dinâmica Molecular (MD); o "Generalized Simulated Annealing" (GSA)[70] e o "Generalized Genetic Algorithm" (GGA)[94]. Tipicamente, o espectro (série) temporal de proteínas apresenta-se com um perfil bastante irregular, que lembra em alguns aspectos os chamados movimentos Brownianos fracionários [95].

Em 2001, Moret e colaboradores [49] aplicaram uma metodologia que combina a técnica GSA com dinâmica molecular, implementada na rotina THOR [96, 97], com o intuito de analisar perfis da hipersuperfície de energia potencial de proteínas em função do tamanho do sistema. Os resultados obtidos para os perfis indicam que os valores locais das energias são distribuídos de forma bastante irregular sobre a hiper-superfície; e que diferentes regiões do espaço de fase apresentavam perfis com uma grande similaridade entre si.

Anteriormente, Lidar e colaboradores [98] realizaram simulações de dinâmica

molecular (programa MOIL), para obter séries temporais de energia potencial de alguns sistemas moleculares, tais como: mioglobina; polialaninas, com $N = 16$ resíduos, entre outros. Basicamente, todos os sistemas investigados foram submetidos à uma temperatura $T = 300K$ e um tempo de simulação no intervalo $10 < t < 25$ picosegundos. O objetivo principal da investigação era, usando um método variacional de análise fractal ("ϵ-variation method"), estudar as propriedades fractais das séries de energia potencial em função do tempo. Para todos os sistemas analisados, os principais resultados obtidos por eles foram: o valor da dimensão fractal (expoente de rugosidade) depende fracamente da temperatura; aumenta com o tempo, mas lentamente, quanto maior for a proteína. Mostraram, também, que a presença de estruturas hélices-α suaviza a rugosidade da séries. Observaram, ainda, indícios de um comportamento universal, ou seja, diferentes sistemas têm sua rugosidade descrita pela mesma dimensão fractal e concluíram que: "a fractalidade das séries não está associada à rugosidade da hiper-superfície de energia potencial, nem às propriedades topológicas da proteína". Confirmar se este cenário sobrevive a uma análise fractal, com uma técnica capaz de vasculhar detalhes finos das séries, tornou-se uma questão a ser esclarecida.

No entanto, uma análise dos mesmos perfis, não em função do tempo, mas em função do número de ângulos diedrais, constatou que estes perfis são objetos multifractais, caracterizados pelos chamados espectros $f(\alpha)$ [49]. Além disso, foi observado que tais espectros são sensíveis ao número de graus de liberdade do sistema, mostrando que a dimensão do espaço de fase influencia a acessibilidade de partes da hiper-superfície, determinando as regiões permitidas e não-permitidas. Este resultado é de fundamental importância numa possível elucidação do Paradoxo de Levinthal, uma vez que existiriam conformações simplesmente inacessíveis ao

sistema, o que diminuiria seu tempo de busca por um mínimo global de energia.

Neste Capítulo, pretendemos apresentar uma abordagem para a caracterização multifractal da rugosidade da hiper-superfície de energia potencial, em função do tempo, para estruturas moleculares chamadas de polialaninas. Neste contexto, as paisagens de energia potencial serão tratadas como séries temporais. Investigaremos de que modo ocorre a visita ao espaço de fases do sistema, ao longo do tempo, e quais são os principais efeitos provocados pelas mudanças na temperatura e no tamanho do sistema, na sua evolução dinâmica. A justificativa, motivação e detalhes sobre técnica utilizada, em nosso trabalho, serão apresentadas em seguida.

4.2 Dinâmica molecular dos sistemas protéicos

Simulações de dinâmica molecular têm sido bastante utilizadas, para estudar o problema do enovelamento de proteínas [99, 100]. Em geral, as simulações envolvem um grande esforço computacional, pois a integração das equações de movimento deve ser feita para um sistema muitas partículas. Quando se trata de sistemas moleculares, é sabido que os mesmos podem adotar um número significativo de configurações, que aumentam com o número de graus de liberdade do sistema. Desta forma, cálculos de dinâmica molecular para sistemas protéicos envolvem, necessariamente, a definição de potenciais efetivos, a partir dos quais a força resultante que atua sobre cada partícula é determinada. Logo, a energia potencial ou conformacional deve representar a contribuição de todos os termos de potencial que a constituem. Esta soma envolve as contribuições de energia potencial entre átomos quimicamente ligados e não-ligados, conforme descreveremos a seguir.

1. **Potencial de ligação**: Oscilações nos comprimentos das ligações entre os

átomos, em relação à sua posição de equilíbrio r_0, devido às flutuações térmicas, são comumente descritas por funções harmônicas, tipo lei de Hooke:

$$V_s = \frac{1}{2}\sum_m K_{s_m}(s_m - r_0)^2. \tag{4.1}$$

Nesse caso, a constante de acoplamento entre os átomos K_{s_m} depende não só de suas massas, mas também da natureza da ligação e, como regra geral, essas constantes são parametrizadas a partir do espectro de vibração das moléculas mais simples, formadas pelos mesmos tipos de átomos.[1].

2. **Potencial dos ângulos entre 3 átomos**: De forma similar ao que acontece com o comprimento das ligações, a vibração dos ângulos, determinados entre três átomos consecutivos, pode, também, ser modelado por meio de potenciais harmônicos:

$$V_\theta = \frac{1}{2}\sum_l K_{\theta_l}(\theta_l - \theta_0)^2, \tag{4.2}$$

onde K_{θ_l} é uma constante de Hooke efetiva da ligação obtida através dos espectros de energias rotacionais da moléculas.

3. **Potencial dos ângulos diedrais impróprios**: Comumente a vibração entre os dois planos formados por um átomo central i, que se encontra ligado a outros três átomos (j,k e l), é descrita também por meio de um potencial harmônico:

$$V_\omega = \frac{1}{2}\sum_s K_{\omega_s}(\omega_s - \omega_0)^2, \tag{4.3}$$

onde ω_s é o ângulo entre o plano formado pelos átomos ijk com o plano

[1]Quanticamente pode-se determinar a separação entre cada nível de energia do espectro vibracional como sendo $\Delta E_{vibra} = \hbar(\frac{\mu}{K_{s_m}})^{1/2}$, onde μ é a massa reduzida do sistema.

formado pelos átomos jkl; e K_{ω_s} é a constante de restituição do movimento oscilatório. Este tipo de potencial é bastante utilizado com o objetivo de manter a estrutura tetraédrica média dos carbonos C_α, ligados aos átomos de hidrogênio, que compõem a proteína.

4. **Potencial dos ângulos diedrais próprios ou torcionais**: Qualitativamente, grande parte das mudanças conformacionais sofridas pelas estruturas protéicas, advém de suas rotações em torno das ligações N-C_α e C_α-C, cujas diferenças energéticas são da ordem das flutuações térmicas. Desta forma, a rotação em torno dessas ligações, pode ser descrita por um potencial simétrico definido por:

$$V_\phi = \frac{1}{2}\sum_n K_{\phi_n}(1 + cos(m\phi_n + \phi_0)), \tag{4.4}$$

onde o parâmetro m é o número de mínimos para a torção de uma ligação química; K_{ϕ_n} é a constante de acoplamento da oscilação e ϕ_0 são ângulos cujos valores são determinados a partir de cálculos quânticos em pequenas moléculas ou de dados experimentais oriundos de espectroscopia de raios-X e ressonância magnética nuclear (NMR). É importante salientar que são os ângulos diedrais C-N-C_α-C(ϕ) e N-C_α-C-N (ψ) quem definem as estruturas secundárias da proteínas.

5. **Potencial de van der Waals**: A interação de van der Waals é uma combinação de dois termos: um repulsivo, de curto-alcance e de origem quântica; e um outro atrativo, de longo-alcance, que resulta da interação entre os momentos de dipolo dos átomos, originados das flutuações nas distribuições de carga de um determinado átomo [2]. O potencial de van der Waals, representa

[2]Essa interação de longo-alcance é também denominada de dispersão de London.

estados ligados com uma distância característica da ordem de 1.2 Å a 2.2 Å, é descrito por:

$$V_{vdw} = \sum_{i<j} \left[\frac{C_{12}(ij)}{{r^{12}}_{ij}} - \frac{C_6(ij)}{{r^6}_{ij}} \right], \tag{4.5}$$

onde os parâmetros de acoplamento $C_{12}(ij)$ e $C_6(ij)$ são obtidos a partir de ajustes experimentais.

6. **Potencial Coulombiano**: Descreve as interações eletrostáticas, numa estrutura protéica cujas origens estão relacionadas a dois efeitos: o primeiro, resulta das deslocalizações eletrônicas nos átomos, provocadas pelas suas diferentes eletronegatividades; enquanto, o segundo é proveniente da interação entre grupos ionizáveis e o meio que circunda a estrutura, que usualmente é a água. Em conjunto, esses ingredientes proporcionam uma distribuição "parcial" de cargas nas proteínas, que dá origem a um grande número de interações dipolares.

   Nos métodos tradicionais de campo de força molecular, as cargas são, frequentemente, obtidas a partir de cálculos *ab initio* de densidade de carga no estado eletrônico fundamental. Uma vez que é formado por um elevado número de interações fracas e de longo-alcance, o potencial eletrostático é fundamental na estabilização final da estrutura proteíca. No que se refere aos solventes, embora, muitas vezes os mesmos sejam tratados como meios contínuos, caracterizados por uma constante dielétrica efetiva ϵ_r, o advento de computadores mais potentes, nos últimos anos, tem possibilitado descrições cada vez mais precisas dos potenciais eletrostáticos, incluindo às vezes a própria representação molecular do solvente. Em geral, o potencial Coulombiano é dado

por:

$$V_{el} = \frac{1}{4\pi}\sum_{i<j}\frac{q_i q_j}{\epsilon_0 \epsilon_r r_{ij}}. \tag{4.6}$$

Em resumo, devemos ressaltar que todos os termos, acima descritos, contribuem, com valores característicos distintos [101], para o potencial efetivo total $V(\{r_i\})$:

$$V(\{r_i\}) = V_s + V_\theta + V_\phi + V_\omega + V_{vdw} + V_{el}, \tag{4.7}$$

que descreve toda a estrutura da macromolécula, em função do vetor posição r_i de cada átomo e de outros elementos efetivos da estrutura[3].

4.3 Séries temporais da energia potencial de polialaninas

Nesta Tese, os cálculos numéricos de dinâmica molecular foram todos realizados com o auxílio de um eficiente código computacional: o programa THOR [96], desenvolvido para investigar estruturas de interesse biológico, como, por exemplo, proteínas. O código inclui em sua arquitetura o campo de força GROMOS [65], para simular as interações atômicas na molécula. Especificamente, simulamos polialaninas com diferentes número de resíduos, em diferentes temperaturas de equilíbrio e conformações iniciais.

Polialaninas são utilizadas como protótipos para estudar o processo de enovelamento de estruturas em conformações hélice- α. No Capítulo 2, mostramos que os momentos de dipolo elétrico, que surgem do desbalanceamento de cargas entre os

[3]Nesta abordagem somente átomos de hidrogênio (H), covalentemente ligados aos átomos de oxigênio (O) ou de nitrogênio (N), são considerados explicitamente, enquanto que os grupos CH_1, CH_2 e CH_3 são considerados como uma unidade atômica.

grupos NH e CO das ligações peptídicas; as ligações tipo ponte de hidrogênio e as interações de van der Waals são ingredientes fundamentais no processo cooperativo responsável pela formação de tais estruturas. Este processo é acelerado com o aumento do número de aminoácidos na estrutura. Desta forma, como assinalado por Shoemaker e colaboradores [102], Moret e colaboradores [103] e Rogers [104], um número crítico mínimo de aminoácidos é necessário, a fim de que sejam observadas tais configurações. Do lado oposto, um número crítico superior, também, é esperado devido à desestabilização provocada pelos efeitos entrópicos.

Resultados experimentais relacionados a pequenos polipetídeos e curtos fragmentos de proteínas, dão suporte ao quadro descrito acima, pois essas moléculas não formam hélices-α em meio aquoso. Os primeiros resultados atribuídos a uma significativa formação destas estruturas em água, próximo à 0°C, foram obtidos por Kim e Baldwin [105], usando fragmentos de ribonuclease A e por Marquesee e colaboradores [106], usando polialaninas contendo 16 resíduos.

Do ponto de vista computacional, diversas abordagens de dinâmica molecular em meio aquoso, foram propostas, com o intuito de analisar a questão da estabilidade dessas conformações. Assim, em trabalhos como os de Doruker e colaboradores [107], verifica-se que polialaninas desenovelam-se numa escala de algumas centenas de ps a uma temperatura de 350K. Numa abordagem similar, mas com um modelo que trata o solvente explicitamente e a uma temperatura de 360K, Hitpold e colaboradores [108] observaram que polipetídeos, baseados em alaninas, com aproximadamente 30 resíduos, estabilizam num tempo da ordem de $30ns$. Finalmente, outro estudo, baseado em uma dinâmica molecular para as coordenadas de torção, Bertsch e colaboradores constataram a formação de hélices-α, para polialaninas com 20 resíduos, em simulações de 0.5ns.

Diversos resultados teóricos e experimentais, bem como os obtidos via simulações computacionais, suportam a idéia de que a acessibilidade da hipersuperfície de energia do sistema, depende de parâmetros como: o número de aminoácidos; a constante dielétrica do meio e a temperatura do banho térmico, onde o mesmo se encontra. Nesse trabalho iremos apresentar e discutir resultados de simulações de dinâmica molecular, utilizando o programa THOR[96], que inclui em sua estrutura o campo de força GROMOS [65], para polialaninas, com um número de resíduos N variando entre 8 e 18 aminoácidos, submetidas a diferentes temperaturas.

Em nossos cálculos numéricos, um idêntico protocolo foi adotado em todos os casos. Partimos de uma temperatura inicial $T_i = 1K$, aquecendo o sistema, a uma taxa de $5K/passo$, até atingir a temperatura de equilíbrio desejada. Três temperaturas de equilíbrio foram consideradas: $T = 275K$, $T = 300K$ e $T = 325K$, num meio contínuo, aproximado por uma constante dielétrica relativa $\epsilon_r = 2$. Os incrementos no tempo da dinâmica foram de $5\,10^{-4}$ ps, e em todas as simulações foram realizados $N_{step} = 5\,10^8$ passos, até alcançarmos um intervalo de tempo da ordem de $25ns$. Na contagem do tempo, descartamos o intervalo associado ao período de termalização do sistema. Portanto, considerando somente os últimos $5ns$ de observação, os resultados que obtivemos, para as séries temporais da energia potencial de polialaninas, são os mostrados nas Figuras 4.1; 4.2; 4.3 e 4.4.

Podemos constatar que em todos os casos analisados, as séries exibem rugosidades típicas das observadas em outros fenômenos complexos descritos por séries temporais auto-afins [109]. Devemos observar ainda que, em cada temperatura, os cálculos foram realizados para valores de N entre 8 e 18 resíduos, de forma que os valores de N utilizados na apresentação das séries: $N = 10, 12, 15, 17$ e 18, no caso $T = 275K$; $N = 10, 13, 15, 17$ e 18, em $T = 300K$ e, finalmente, $N = 10, 14, 15, 17$ e

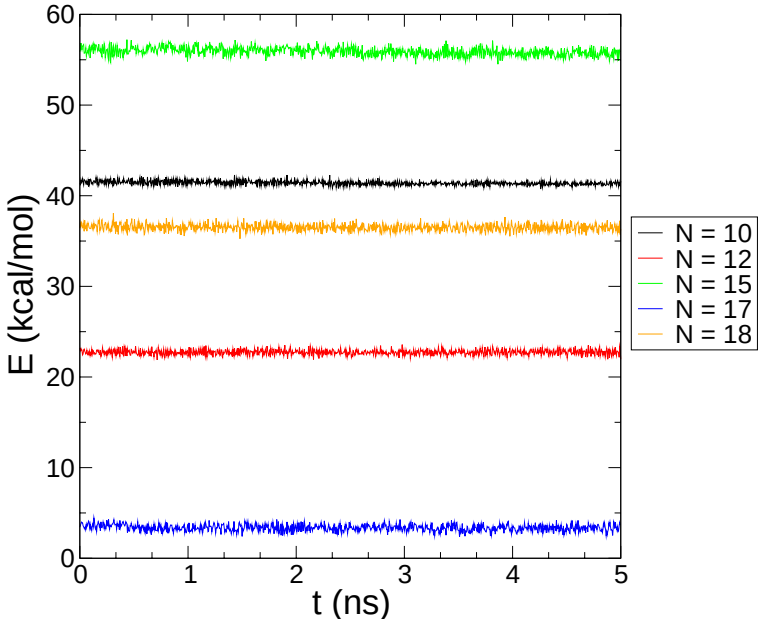

Figura 4.1: Séries temporais da energia potencial de polialaninas com diferentes números de resíduos: N=10 (preto), N=12 (vermelho), N=15(verde), N=17 (azul) e N=18 (laranja). Em todos os casos a temperatura final de termalização é $T = 275K$.

18, para $T = 325K$; ilustram adequadamente o caráter rugoso de todas as séries analisadas. Em particular, a Figura 4.4 exibe detalhes, numa escala de tempo reduzida, da rugosidade embutida nas séries.

Na sequência, discutiremos os resultados relacionados ao estudo que fizemos sobre o comportamento da energia potencial, em função do número de resíduos, considerando uma temperatura T fixa. Veremos que os resultados mostram importantes aspectos relacionados à evolução dinâmica das proteínas.

A Figura 4.5 representa a energia potencial E de polialaninas, em função do número de resíduos N, em $T = 275$. Observe que, inicialmente, os valores de energia E crescem com o aumento de N, no intervalo entre $8 - 10$ resíduos. No intervalo entre $11-12$, os valores de E diminuem, alcançando um valor mais baixo em $N = 12$.

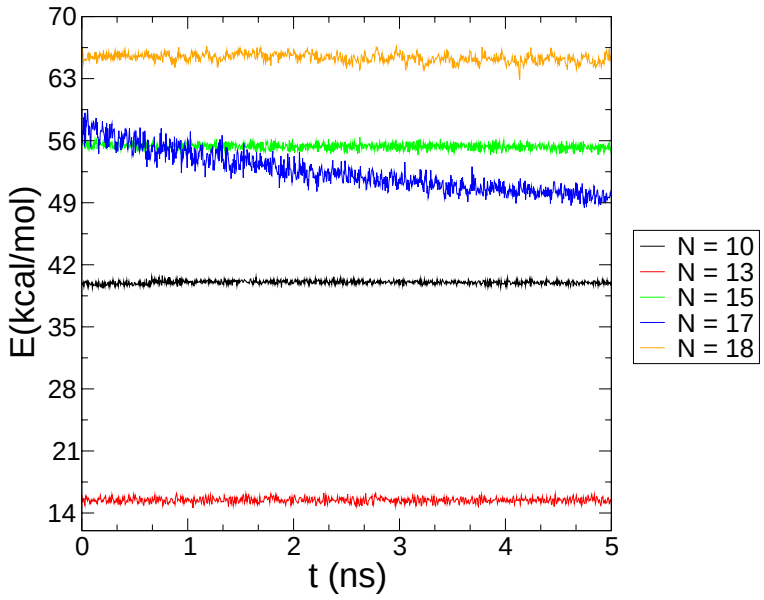

Figura 4.2: Séries temporais da energia potencial de polialaninas com diferentes números de resíduos: N=10 (preto), N=13 (vermelho), N=15 (verde), N=17 (azul) e N=18 (laranja). Em todos os casos a temperatura final de termalização é $T = 300K$.

Em seguida, para N entre $13 - 15$, novamente a energia aumenta com os valores crescentes de N, mas volta a diminuir quando $N = 16$ e em $N = 17$ atingindo o seu valor mínimo. Finalmente, quando $N = 18$, a energia volta a crescer. Em geral, espera-se que energia potencial aumente, em relação a um mínimo global, quando as cadeias protéicas se tornam mais longas [99, 71].

A Figura 4.6 refere-se à energia potencial E em função de N, quando a temperatura é fixada em $T = 300K$. Podemos observar qua a energia apresenta um comportamento similar ao do caso anterior, mas a primeira tendência de crescimento de E ocorre no intervalo entre $8 - 11$ resíduos. Em seguida, temos duas quedas sucessivas em $N = 12$ e $N = 13$. Para o intervalo compreendido entre $14 - 16$, a energia volta aumentar, diminui em $N = 17$ e depois, muda novamente o

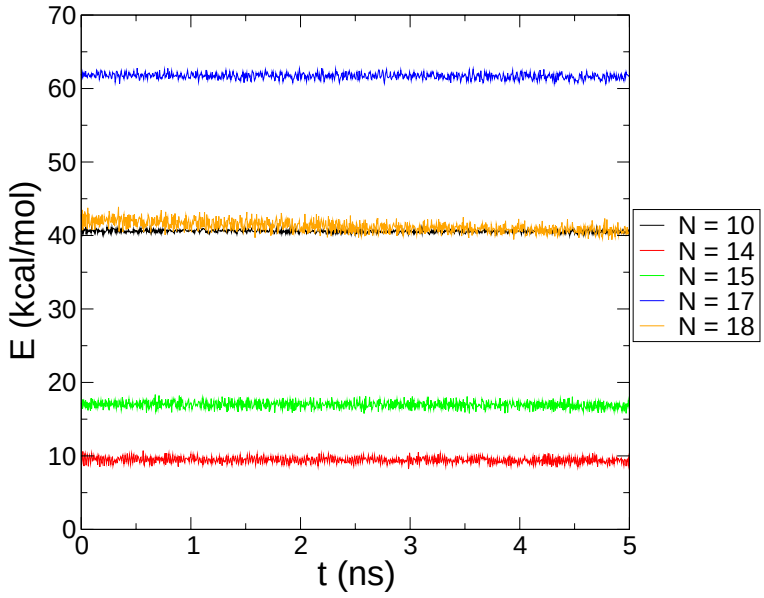

Figura 4.3: Séries temporais da energia potencial de polialaninas com diferentes números de resíduos: N=10 (preto), N=14 (vermelho), N=15 (verde), N=17 (azul) e N=18 (laranja). Em todos os casos a temperatura final de termalização é $T = 325K$.

comportamento, ou seja, atinge o seu maior valor em $N = 18$.

Os resultados encontrados, em $T = 325K$, para a energia E em função de N, são mostrados na Figura 4.7. Neste caso, vemos que o aumento de temperatura amplia o intervalo de crescimento inicial da energia, para a região entre $8 - 15$ resíduos; o valor mínimo acontece em $N = 14$; e, entre $16 - 18$ resíduos ocorrem dois aumentos sucessivos, seguido de uma queda para $N = 18$.

Para compreendermos essa dependência da energia potencial em relação ao número de resíduos, uma vez fixada a temperatura, devemos associá-la ao processo dinâmico de formação de estruturas intermediárias que ocorre ao longo do tempo. Dois cenários confirmam esta relação: o primeiro, baseado em teorias sobre o enovelamento de proteínas, inspiradas na análise de hiper-superfícies de energia confor-

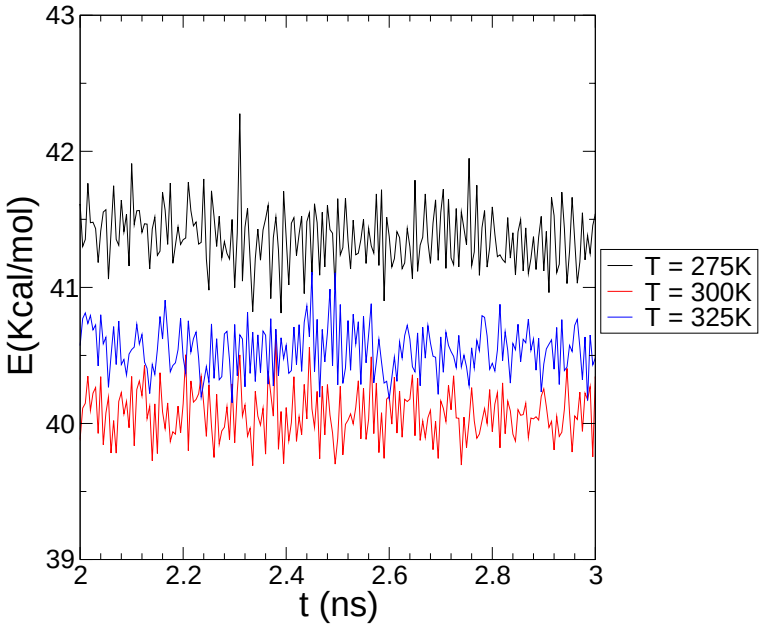

Figura 4.4: Detalhes da região entre $2ns$ e $3ns$, para séries temporais com N=10 resíduos e diferentes temperaturas de termalização: $T = 275K$ (preto), $T = 300K$ (vermelho) e $T = 325K$ (azul).

macional, assegura que, em média, a energia deve diminuir, quanto mais estruturas intermediárias vão se formando ao longo do processo de busca da estrutura nativa [71, 110]; o segundo, ampara-se na constatação de Moret e colaboradores [103] de que as polialaninas adquirem estáveis estruturas secundárias, especificamente, hélices-α, quando o número de resíduos N for maior ou igual a 13.

Em princípio, podemos admitir, que os resultados que obtivemos estão condizentes com as observações acima citadas, pois se tomarmos o caso $T = 300K$, o menor valor de energia acontece extamente em $N = 13$ (ver Figura 4.6). Provavelmente, o crescimento inicial da energia, para N entre $8 - 11$, pode ser atribuído ao fato de que, neste intervalo, a nucleação de estruturas secundárias ainda é pequena e as proteínas têm grande mobilidade. Então, as flutuações térmicas e o aumento

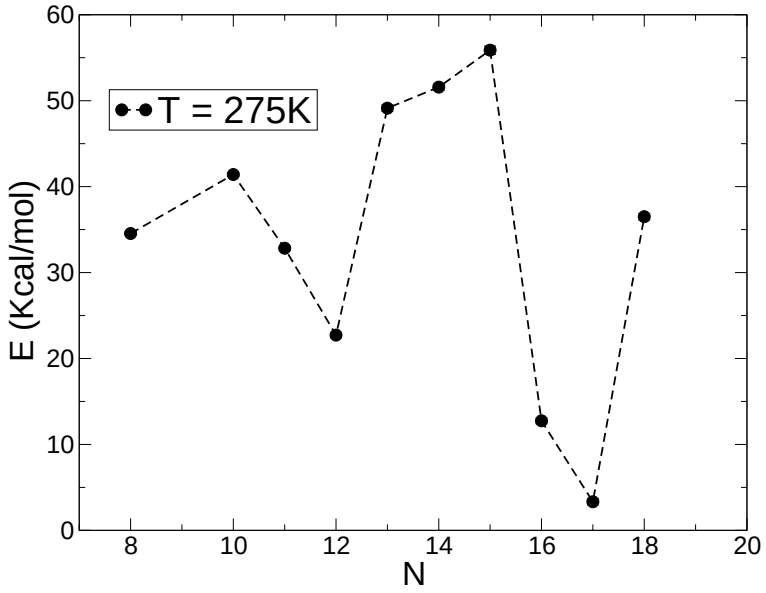

Figura 4.5: Energia potencial das polialaninas em função do número de resíduos, em $T = 275K$.

no número de resíduos favorecem o aumento da energia conformacional.

Por outro lado, nossos resultados sugerem que, entre $14 - 18$ resíduos, o aumento da energia E, está associado ao surgimento de novos e crescentes domínios de estruturas secundárias, do tipo hélice-α, em virtude do aumento no número de pontes de hidrogênio, ao longo da cadeia protéica, que são as interações responsáveis pelo aumento na estabilidade do sistema. Em $N = 17$, a polialanina alcança uma nova fase de estabilidade, enquanto, em $N = 18$, a energia volta crescer, devido, ao que parece, à combinação de dois efeitos: o aumento no número de resíduos e do número de hélices-α.

No caso $T = 275K$, ou seja, quando diminuímos a temperatura, constatamos que a energia apresenta dois mínimos: em $N = 12$ e em $N = 17$ (ver Figura 4.5). Comparando com a situação, acima descrita, para $T = 300K$, isto parece indicar

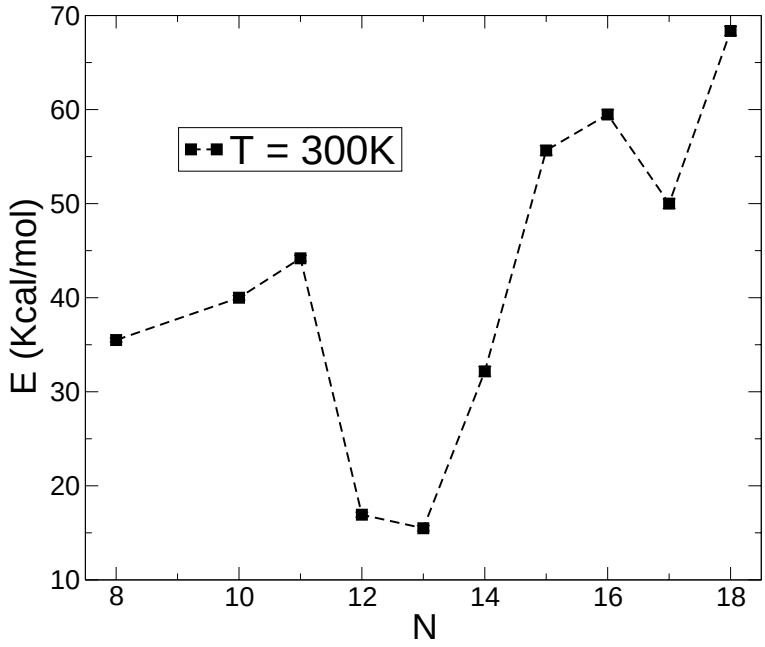

Figura 4.6: Energia potencial das polialaninas em função do número de resíduos, em $T = 300K$.

que, numa temperatura mais baixa, o número crítico de resíduos, para estabilizar as estruturas em hélice-α deve diminuir, que no, presente caso, seria igual a 12. Em relação ao comportamento de E, para os demais valores de N, podemos considerar que se deve, também, ao aumento no número de resíduos e, consequentemente, do número de hélices-α.

Finalmente, analisando o que acontece quando aumentamos a temperatura, ou seja, para $T = 325K$, podemos constatar na Figura 4.7 a existência de dois mínimos de energia, que surgem em $N = 14$ e $N = 18$. Logo, quando a temperatura aumenta, apenas um resíduo a mais na estrutura é necessário, para iniciar o processo de formação de estruturas secundárias ($N = 14$). A proteína apresenta uma grande estabilidade em $N = 18$; e para os outros valores de N, o comportamento da energia

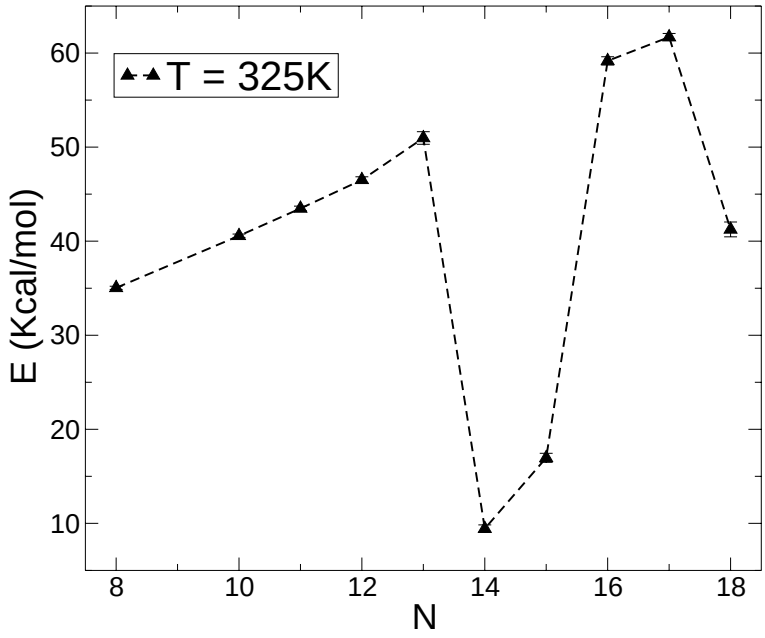

Figura 4.7: Energia potencial das polialaninas em função do número de resíduos, em $T = 325K$.

é análogo aos dos casos anteriores.

Portanto, levando em conta de que a rugosidade das séries de energia potencial representam paisagens da hiper-superfície de energia conformacional de proteínas, uma caracterização cuidadosa das flutuações estatísticas presentes nas séries se faz necessária. O objetivo é o de captar, nos dados contidos nas séries, informações sobre o comportamento dinâmico do sistema. Para realizarmos essa tarefa aplicaremos a metodologia MF-DFA, conforme descreveremos na seção seguinte.

4.4 O método MF-DFA

Muitos sistemas físicos e biológicos não apresentam uma escala de compri-

mento ou de tempo característico, de forma que algumas de suas grandezas exibem correlações espaciais ou temporais de longo alcance, descritas por leis de potência [111, 112]. De forma geral, o estudo da dinâmica desses sistemas, envolve a caracterização estatística das correlações de longo-alcance, presentes nas séries temporais, associadas a esses sistemas. Esta tarefa encontra algumas dificuldades, se as séries apresentarem propriedades estatísticas heterogêneas, ou seja, flutuações locais próprias ou induzidas, que introduzem tendências nas mesmas.

Portanto, precisamos usar técnicas que levem em conta essa possibilidade, devido ao fato de que as tendências podem provocar comportamentos "espúrios", que afetam as correlações presentes nas séries. Nos últimos anos, as principais técnicas utilizadas para eliminar tendências em séries temporais são: a técnica de *Análise de Flutações Destendenciada* DFA (*Detrended Fluctuation Analysis*) e o método da transformada de ondaletas (*Wavelets Transform*) [113, 114]. Ambas as técnicas se baseiam em ajustes polinomiais locais, para tratar adequadamente as tendências locais em diferentes segmentos da série. Em particular, DFA tem sido empregada com eficiência em problemas de diversas áreas como, por exemplo: sequências de DNA[115]; análise de batimentos cardíacos [116]; economia [1]; meteorologia [117]; astrofísica [118], entre outros.

Em geral, a opção de aplicar DFA nesses estudos, se deve à sua facilidade de implementação, pois trata-se de uma ferramenta que permite analisar, em séries temporais auto-afins, estacionárias, o papel das tendências, bem como realizar de maneira eficiente uma estimativa das correlações de longo-alcance, através de um único parâmetro: o expoente de escala α ou expoente H de Hurst. Devemos salientar que o tipo de correlação nas séries estacionárias depende do valor encontrado para o expoente H, pois quando a $H = 0.5$ o sinal é descorrelacionado (ruído branco

ou Gaussiano), enquanto para a $H < 0.5$ temos anti-correlação (anti-persistência) e para a $H > 0.5$ dizemos que existe correlação (persistência)[95].

Contudo, a técnica de DFA não fornece resultados satisfatórios quando confrontada com séries temporais não estacionárias, ou seja, séries que são afetadas por tendências locais, pois neste caso, diferentes partes da série exibem distintos comportamentos de escala. Logo, essas séries não são descritas completamente por um único expoente H, uma vez que diferentes segmentos da série apresentam flutuações caracterizadas por distintos valores de H. Neste caso, o formalismo adequado para determinar a distribuição de valores dos expoentes de escala é a análise multifractal.

Nos últimos anos, têm crescido o número de trabalhos voltados para a caracterização de multifractalidade em séries temporais, não estacionárias, constituídas, principalmente, de dados experimentais [119]. Embora, nesses problemas, possamos aplicar análise de ondaletas [114]; nesta Tese, utilizaremos a generalização da técnica DFA proposta por J.W. Kanterlhardt e colaboradores [120], para investigar séries temporais de energia potencial de proteínas.

Esta generalização denominada MF-DFA, *Multifractal Detrended Fluctuation Analysis ou Análise Multifractal de Flutuações Destendenciadas*, tem sido aplicada com sucesso em: sequências de DNA [121], meteorologia [122], sismologia [123], entre outras. Além disso, a mesma é, também, de fácil implementação e tem demonstrado excelente perfomance na obtenção de resultados relacionados, tanto a dados artificiais, quanto aos reais, em comparação ao desempenho da análise de ondaletas, para os mesmos sistemas [119]. Na sequência iremos descrever, brevemente, o método MF-DFA.

A aplicação da técnica MF-DFA consiste dos seguintes passos:

1. Considere sobre um suporte compacto, uma série $u(i)$, $i = \{1, ..., N_{max}\}$, onde

N_{max} é o comprimento da série, para determinarmos o seu perfil (série integrada), ou seja,

$$y(i) = \sum_{k=1}^{i} [u(k) - < u >], \tag{4.8}$$

onde $< u >$ é a média tomada sobre todos os valores da série original $u(i)$;

2. Divida o perfil em N_s caixas (segmentos disjuntos) de igual tamanho $\underline{s}$. A partição da série deve ser feita duas vezes: do início ao final da série e no sentido inverso, gerando um total de $2N_s$ segmentos. Cada caixa $\underline{\nu}$ tem sua própria tendência local, que pode ser aproximada através de um ajuste polinomial de ordem $\underline{m}$, que será subtraído dos dados, representado pela variável de $y_s(i)$;

3. Calcule a tendência local $\underline{y_s(i)}$ em cada um dos $2N_s$ segmentos de tamanho $\underline{s}$. Em seguida, determine o desvio quadrático médio ou variância da flutuação associada a cada segmento:

$$F^2(s,\nu) \equiv \frac{1}{s}\sum_{i=1}^{s}\{y[(\nu-1)s+i] - y_\nu(i)\}^2 \quad \nu = \{1, ..., N_s\}; \tag{4.9}$$

4. Agora, faça uma média sobre todos os $2N_s$ segmentos, para obter a função de flutuação de ordem q:

$$F_q(s) \equiv \left\{ \frac{1}{2N_s} \sum_{\nu=1}^{2N_s} [F^2(s,\nu)]^{q/2} \right\}^{1/q}; \tag{4.10}$$

onde, em geral, a variável q assume valores reais.

5. Finalmente, analisando gráficos log-log de $F_q(s)$ em função de s, para cada va-

lor de q, podemos determinar o comportamento de escala da função flutuação. Se as séries $u(i)$ apresentarem correlações de longo-alcance, então para valores crescentes de s, as funções $F_q(s)$ também crescem, seguindo uma lei de potência do tipo:

$$F_q(s) \sim s^{h(q)}, \tag{4.11}$$

onde $h(q)$ é denominado expoente de Hurst generalizado.

Um importante parâmetro desse método é a ordem $m = 1, 2, 3...$ do polinômio usado no ajuste dos dados em cada segmento. A escolha de m corresponde a diferentes ordens de DFA, que diferem entre si na capacidade de eliminar as tendências. Por exemplo, se $m = 0$, somente tendências constantes podem ser eliminadas nas séries; se $m = 1$, tendências constantes e lineares são eliminadas; e assim por diante. O resultado obtido com MF-DFA é uma família de expoentes $h(q)$, que para uma genuína série multifratal, formam uma função descrescente de q, enquanto, se o sinal for monofractal $h(q)$ independe de q. Em particular, quando $q = 2$, $h(2) = H$ é o clássico expoente de Hurst. Por outro lado, se $q < 0$, $h(q)$ captura as propriedades das pequenas flutuações, enquanto para $q > 0$ as grandes flutuações são dominantes. Temos ainda que, existe uma relação simples entre o expoente $h(q)$ e o expoente de escala multifractal $\tau(q)$, definida pelo formalismo multifractal [120]:

$$\tau(q) \equiv qh(q) - 1. \tag{4.12}$$

A função $\tau(q)$ é uma das mais utilizadas representações, para espectros multifractais relacionados a medidas em séries temporais. Outras formas de ca-

racterizar singularidades em séries multifractais são: o espectro de dimensões generalizadas,

$$D(q) \equiv \frac{\tau(q)}{q-1}, \tag{4.13}$$

e o espectro de singularidades,

$$f(\alpha) = q[\alpha - h(q)] + 1, \tag{4.14}$$

tal que,

$$\alpha = h(q) + qh'(q). \tag{4.15}$$

Na próxima seção, iremos apresentar os principais potenciais envolvidos nas simulações de dinâmica molecular, que realizamos para obtermos as séries temporais da energia potencial de polialaninas.

4.5 Caracterização multifractal das séries de energia potencial

Para estimarmos os valores do expoente $h(q)$ e, consequentemente, do espectro $\tau(q)$, a partir das séries temporais obtidas e discutidas na seção anterior, aplicamos a técnica MF-DFA (ver Seção 4.2). Inicialmente, calculamos as funções flutuações $F_q(s)$ em função do número de resíduos e da temperatura. Em todos os casos, utilizamos séries de comprimento $N_s = 20000$ pontos; escolhemos o parâmtro q no intervalo $-10 < q < 10$, com passo $\Delta q = 0.5$; e, para encontramos o melhor

ajuste linear dos dados, variamos, tanto a ordem m do polinômio de ajuste das tendências, quanto o valor da maior e da menor partição s da série.

A Figura 4.8 representa linearizações (gráficos log-log) da função flutuação $F_q(s)$ em função da escala s e do parâmetro q, para as séries mostradas na Figura 4.1, ou seja, com $N = 10, 12, 15, 17$ e 18; e $T = 275K$. Em média foram escolhidos os valores de escala no intervalo $20 < s < 80$ e as tendências foram aproximadas por um polinômio de ordem $m = 1$. Conforme podemos observar, as estimativas obtidas para os ajustes lineares dos dados atendem satisfatóriamente ao comportamento de escala previsto pela Equação 4.11. Na Figura 4.9(a) mostramos os expoentes de $h(q)$ (valores das inclinações das retas ajustadas aos dados) em função de q; enquanto a Figura 4.9(b) exibe o correspondente espectro multifractal, $\tau(q)$ em função de q, obtido com os respectivos valores de $h(q)$.

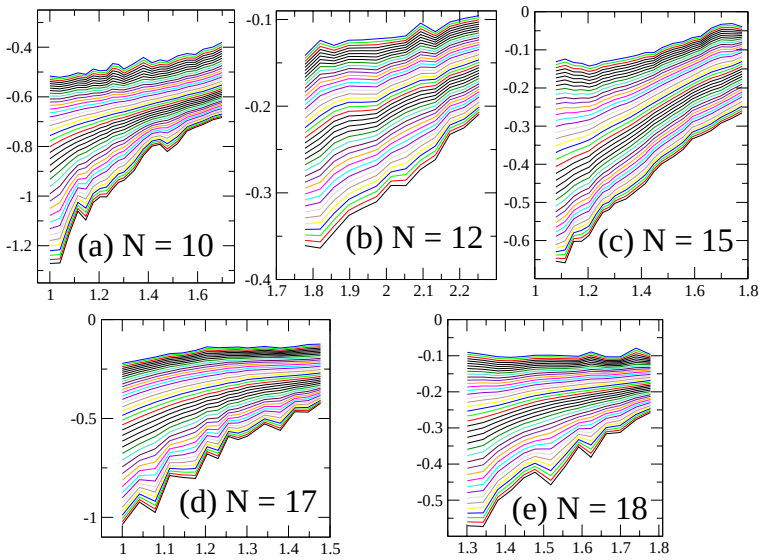

Figura 4.8: Comportamento de escala da flutuação $F_q(s)$ em função da escala s, para as séries temporais de energia exibidas na Figura 4.1 ($T = 275K$).

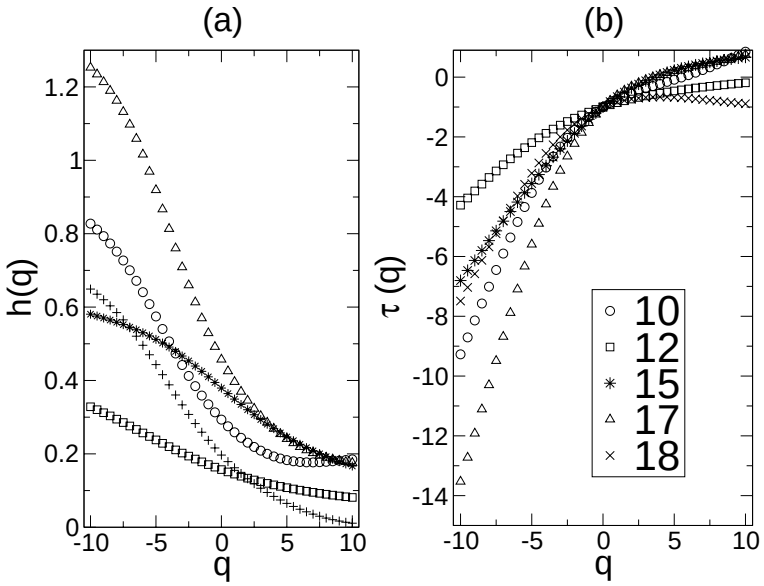

Figura 4.9: (a) Expoentes de Hurst generalizados $h(q)$ em função de q. (b) Espectro multifractal $\tau(q)$ em função de q. Os dados referem-se às polialaninas cujas $F_q(s)$ são mostradas na Figura 4.8.

No caso das séries exibidas na Figura 4.2, ou seja, com $N = 10, 13, 15, 17$ e 18; e $T = 300K$, usamos um intervalo de escala $20 < s < 100$ e um polinômio de ajuste de ordem $m = 4$. Os resultados encontrados para a função flutuação $F_q(s)$ em função da escala s e do parâmetro q, mostrados na Figura 4.10, seguem um típico comportamento de escala (ver Equação 4.11). Os valores de $h(q)$, associados aos dados da Figura 4.10, são os contidos na Figura 4.11(a), enquanto o correspondente espectro $\tau(q)$ é mostrado na Figura 4.11(b). Finalmente, para as séries apresentadas na Figura 4.3, ou seja, com $N = 10, 14, 15, 17$ e 18; e $T = 325K$; escolhemos $12 < s < 50$ e um polinômio de ajuste de ordem $m = 1$. Os resultados encontrados para $F_q(s)$; $h(q)$ e $\tau(q)$ são apresentados nas Figuras 4.12; 4.13(a) e 4.13(b), respectivamente.

De uma maneira geral, podemos afirmar que o conjunto de resultados apre-

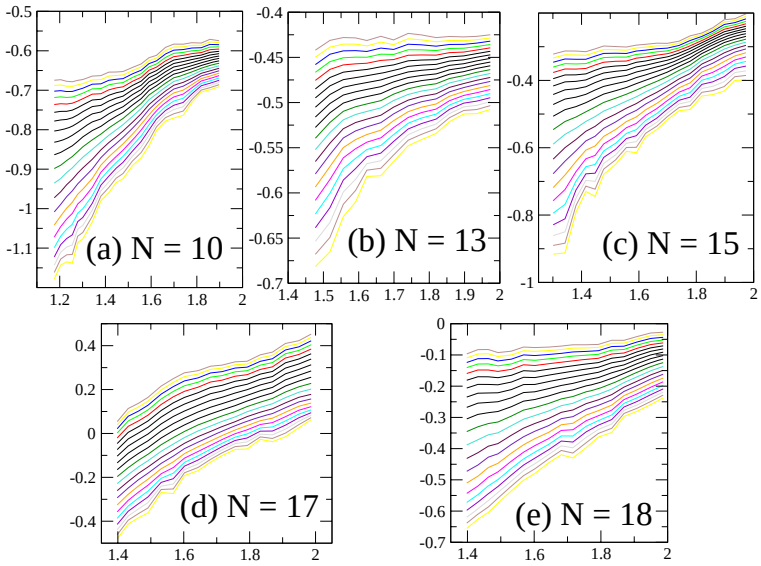

Figura 4.10: Comportamento de escala da flutuação $F_q(s)$ em função da escala s, para as séries temporais de energia exibidas na Figura 4.2 ($T = 300K$).

sentados acima, indicam que as séries temporais investigadas exibem um típico comportamento multifractal, que depende do número de resíduos N na polialanina e da temperatura T de equilíbrio da mesma. No primeiro caso, $T = 275K$, a função $h(q)$ descresce com o aumento de q e o espectro $\tau(q)$ não é uma função linear de q (ver Figuras 4.9(a) e 4.9(b)). Em particular, para $N = 12$, vemos que a multifractalidade é bastante atenuada. Além disso, podem ser observados diferentes regimes de correlação: para $N = 10$, a série é totalmente correlacionada; enquanto para $N = 15, 17$ e 18, temos um regime misto, ou seja, anti-correlação quando $q > 0$ e correlação, para alguns valores de $q < 0$. A série identificada com $N = 17$ apresenta regiões de correlações maiores do que as séries relacionadas a $N = 15$ e $N = 18$.

No caso $N = 12$, número crítico de resíduos associado à formação de hélices-α em $T = 275K$, a série é totalmente anti-correlacionada. Desde que para este valor

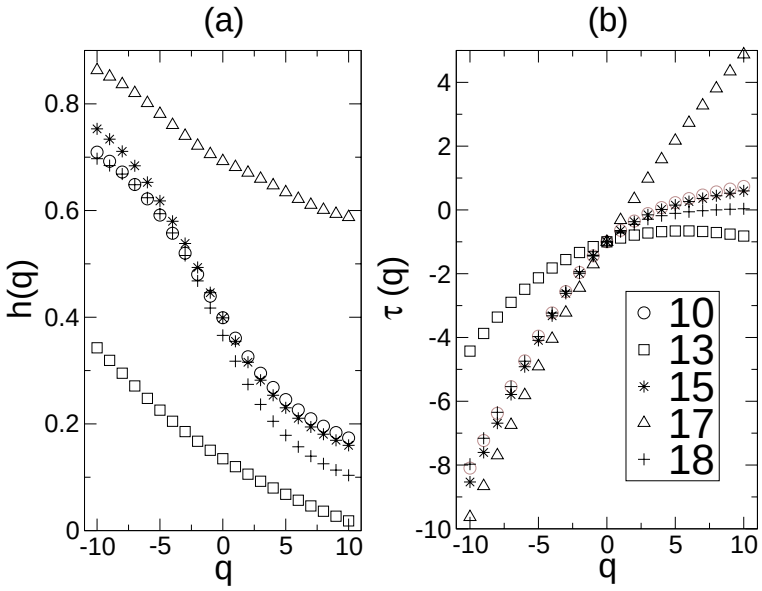

Figura 4.11: (a) Expoentes de Hurst generalizados $h(q)$ em função de q. (b) Espectro multifractal $\tau(q)$ em função de q. Os dados referem-se às polialaninas cujas $F_q(s)$ são mostradas na Figura 4.10.

de N, a energia potencial é um mínimo secundário (ver Figura 4.5), então devemos admitir que a nucleação de estruturas secundárias altera a dinâmica do sistema para um regime anti-correlacionado, superando, portanto, a tendência de crescimento da energia induzida pela agitação térmica e o aumento do número de resíduos.

Em relação a $N = 17$ (mínimo global), a presença de dois regimes de correlação pode ser atribuída: a um aumento da rugosidade da série, devido um número maior de visitas ao espaço de fases (tempos de observação longos) e ao papel das hélices-α, na estabilidade da estrutura protéica, que dificulta a mobilidade da proteína e suaviza a rugosidade da hiper-superfície de energia do sistema. No que se refere a $N = 15$ e $N = 18$, a tendência de crescimento na energia é amortecida pelo forte crescimento dos domínios de hélices-α, principalmente, quando $N = 18$,

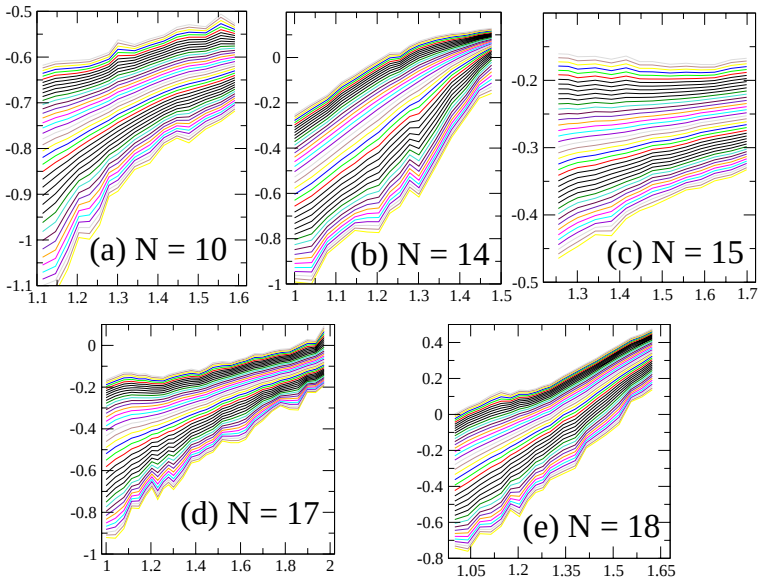

Figura 4.12: Comportamento de escala da flutuação $F_q(s)$ em função da escala s, para as séries temporais de energia exibidas na Figura 4.3 ($T = 325K$).

que deve ser o sistema com o maior número de regiões em conformações hélice-α.

Um quadro análogo ao descrito acima, acontece no segundo caso, $T = 300K$ (ver Figuras 4.11(a) e 4.11(b)) . Destaca-se sobretudo o comportamento das funções $h(q)$ e $\tau(q)$, referentes à $N = 12$ e 13, que possuem valores de energia muito próximos, com $N = 13$ representando o mínimo global (ver Figura 4.6). Para esses valores de N, as séries sinalizam com uma dinâmica influenciada por uma intensa nucleação de estruturas secundárias que compete com o crescimento da agitação térmica, provocada pelo pequeno aumento de temperatura. No mais, em $T = 300K$, todos os espectros são multifractais.

Finalmente, para o terceiro caso, $T = 325K$, encontramos séries com um regime de correlação mista para $N = 13$ e 17; anti-correlacionadas para $N = 15$ e correlacionadas para $N = 14$ e 18, conforme as Figuras 4.13(a) e 4.13(b). Novamente

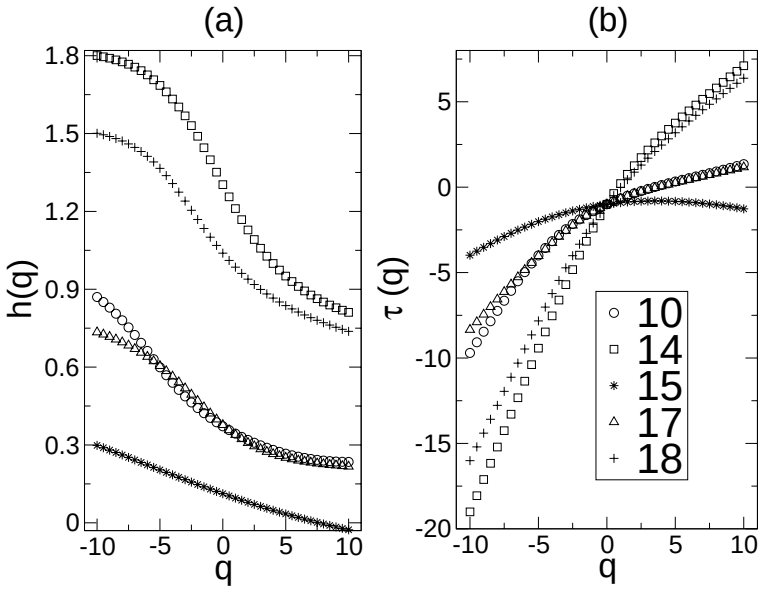

Figura 4.13: (a) Expoentes de Hurst generalizados $h(q)$ em função de q. (b) Espectro multifractal $\tau(q)$ em função de q. Os dados referem-se às polialaninas cujas $F_q(s)$ são mostradas na Figura 4.12.

atribuímos ao fraco comportamento multifractal para $N = 15$, ao relevante papel exercido pelas hélices-α na dinâmica do sistema, pois o processo de formação dessas estruturas aumenta nas cadeia mais longas. Por esta razão, $N = 18$ deve ser a estrutura mais estável, quando $T = 325K$.

Capítulo 5

Conclusões e perspectivas

"Sabido é que todo o efeito tem sua causa, e esta é uma universal verdade, porém, não é possível evitar alguns erros de juízo, ou de simples identificação, pois acontece considerarmos que este efeito provém daquela causa, quando afinal ela foi outra, muito fora do alcance do entendimento que temos e da ciência que julgávamos ter."

José Saramago - A jangada de pedra

Estudamos, nesta Tese, aspectos espaciais e temporais do problema do enovelamento protéico. Procuramos não apenas esclarecer os ingredientes fundamentais por trás de diversos fatos experimentais, mas também tivemos o objetivo de sugerir modelos e abordagens mais simplificados que permitissem abordar os mecanismos do enovelamento protéico, seja do ponto de vista puramente estrutural ou dinâmico. Desta forma, no Capítulo 3 através de um modelo de caminhante aleatório, obtivemos a relação de escala entre o raio de giração R_g e o número de resíduos na cadeia

principal N,

$$R_g \sim N^{\nu}, \tag{5.1}$$

para diversos valores da fração f, de motivos tipo hélice-α e/ou folha-β, e do desvio padrão da distribuição de probabilidades angular δ. Tal estudo permitiu observar uma grande variedade de comportamentos, para expoente ν, desde aquele totalmente difusivo, onde $\nu = 0.5$, correspondendo ao limite em que $\delta \to \pi$; passando pelo valor mínimo obtido por meio da teoria do colapso polimérico [124], onde $\nu \sim 1/3$, correspondendo ao limite $\delta \to 0.15\pi$ e pelo predito pela teoria de Flory [82], onde $\nu = 0.60$; chegando finalmente ao extremo de uma cadeia extremamente rígida, onde $\nu = 1.00$, correspondendo ao limite de aleatoriedade nula ($\delta \to 0$).

Concluímos que a porcentagem $f \cong 60\%$ de motivos específicos e largura $\delta \cong 0.15\pi$ são os parâmetros de nosso modelo que determinam o expoente ν mais próximo àquele verificado experimentalmente $\nu_{exp} = 0.40 \pm 0.02$; seja para o caso hélice-α $\nu_{\alpha} = 0.401 \pm 0.002$, folha-$\beta$ $\nu_{\beta} = 0.417 \pm 0.002$ ou para aquele em que temos uma mistura equitativa dos dois motivos, anteriormente especificados, $\nu_{mix} = 0.409 \pm 0.002$. Determinamos, também, o papel da fração f no grau de compactação da estrutura, definida pela grandeza γ, concluindo que o valor $f \cong 0.60$ também é aquele que maximiza a razão entre o raio de giração do sistema R_g e a separação entre os dois elementos mais distantes na estrutura D_{max}.

Ainda, a respeito da fração f, é importante ressaltar sua relevância na distribuição do número de coordenação z_c e da energia entre os contatos E, calculada em nosso modelo, a partir da energia de interação entre aminoácidos, como sugerida por diversos trabalhos na literatura [73]. Neste contexto, novamente, o valor $f = 0.60$ é aquele que melhor descreve os resultados experimentais para cadeias

protéicas globulares extraídas do PDB.

Estes resultados, abrem caminho para novas investigações acerca do papel da estocasticidade da distribuição angular, na determinação de estruturas mais compactas e do papel da energia entre os contatos, na estabilização destas estruturas. Indicando como perspectiva imediata a necessidade de um estudo sobre a dependência da energia E entre os contatos e do parâmetro de compactação γ, com a largura da distribuição angular de probabilidades δ.

No Capítulo 4 desta Tese, estudamos as propriedades multifractais de séries temporais da energia potencial de polialaninas. Foram analisadas cadeias protéicas, com diferentes números de resíduos, em três temperaturas de equilíbrio. A investigação foi realizada com uma abordagem de dinâmica molecular combinada com uma técnica de análise estatística, que nos permitiu caracterizar a rugosidade associada às correlações temporais existentes entre as variáveis dinâmicas das séries.

Neste trabalho, as séries temporais de energia potencial foram obtidas de simulações de dinâmica molecular usando o programa THOR [96] combinado com o campo de força GROMOS [65]. A metodologia MF-DFA (ver Seção 4.2) foi aplicada na análise da rugosidade das séries. Típicamente, nossos resultados mostram que todas as séries analisadas exibem um comportamento multifractal que depende, tanto do número de resíduos N, quanto da temperatura T do sistema. Além disso, as propriedades multifractais das séries, representadas pelos espectros $\tau(q)$ ou pelos expoentes de Hurst generalizados $h(q)$, revelam aspectos importantes sobre a evolução temporal do sistema.

Foi constatado, por exemplo, que quando o número de resíduos se aproxima do número crítico de resíduos N_c, associado à formação de uma quantidade significativa de estruturas secundárias (hélices-α), o regime de correlação temporal do

sistema é alterado. Em $N_c = 12$ e $T = 275K$, a série é totalmente anti-correlacionada e de maneira intensa no domínio das grandes flutuações ($q > 0$), mas $\tau(q)$ depende fracamente de q, conforme Figura 4.9. No caso, $N_c = 13$ e $T = 300K$, o comportamento das correlações é, também, totalmente anti-correlacionado, mas o espectro $\tau(q)$ é genuínamente multifractal e a rugosidade é mais acentuada na região das pequenas flutuações ($q < 0$), como pode ser visto na Figura 4.11. Finalmente, para $N_c = 14$ e $T = 325K$, o espectro $\tau(q)$ descreve uma série totalmente correlacionada, com um relativo predomínio das pequenas flutuações ($q < 0$), como exibido na Figura 4.13. Para os demais valores de N, nossos resultados mostram que os dois regimes de correlação estão presentes nas séries.

Portanto, a metodologia MF-DFA aplicada a séries temporais de energia potencial de proteínas, especificamente polialaninas, permitiu revelar aspectos importantes sobre a riqueza e complexidade, relacionada a evolução temporal desses sistemas, em busca de seu estado nativo. De fato, em função do número de resíduos e da temperatura, mostramos que a trajetória da proteína é guiada, principalmente, pela influência das estruturas secundárias, que vão sendo formadas ao longo do tempo de simulação, vasculhando a hiper-superfície de energia conformacional em diferentes escalas de tempo. Em decorrência disso, as séries de energia possuem correlações temporais de longo-alcance multifractais.

Nossos resultados, sugerem, que embora alguns resultados obtidos, por Lidar e colaboradores [98], como, por exemplo, a influência do tempo e das hélices-α na rugosidade das séries sejam pertinentes; as demais conclusões não se confirmam, em relação aos nossos dados, pois o tempo de simulação utilizado por aqueles autores é muito menor do que o utilizado no presente estudo, ou seja, insuficiente para uma adequada formação de estruturas secundárias; e a técnica de análise fractal, que

não trata adequadamente a existência de tendências nas séries, não possibilitou aos citados autores captar detalhes finos dos espectros.

Por outro lado, se considerarmos que um sistema que possui uma paisagem de energia complexa, rugosa, com muitos mínimos e vales separados por barreiras de diversas alturas, deve, ao visitar dinamicamente o seu espaço de fases, experimentar uma variedade de escalas de tempo; podendo neste processo tanto oscilar, como saltar de um vale para outro, então, nossos resultados indicam que este comportamento deve ser a origem da observada multifractalidade, no tempo, da hiper-superfície de energia potencial.

Outra questão, sobre os nossos resultados, é de que os mesmos parecem concordar com as evidências, atribuídas a estudos de dinâmica molecular, que sugerem que o enovelamento de proteínas é iniciado com fragmentos específicos (estruturas intermediárias) que rapidamente adotam conformações próximas das encontradas para as estruturas nativas das mesmas [125].

Um outro aspecto dos nossos resultados dão suporte a uma explicação alternativa sobre o chamado Paradoxo de Levinthal. Em um trabalho anterior, Moret e colaboradores [49] fizeram uma análise dos espectros (perfis) de energia potencial, similar à nossa, não em função do tempo, mas em função do número de ângulos diedrais ϕ e ψ, encontrando que estes perfis são objetos multifractais, característica tipicamente descrita pelos chamados espectros $f(\alpha)$. Observaram também que estes espectros $f(\alpha)$ eram sensíveis ao número de graus de liberdade do sistema, mostrando que a dimensão do espaço de fase do problema influencia a acessibilidade de partes da hiper-superfície de energia potencial, ou seja, os resultados encontrados mostram que as proteínas adotam conformações no espaço de fase somente nas regiões permitidas do espectro $f(\alpha)$. Logo, a hiper-superfície de energia potencial

apresenta regiões permitidas e não-permitidas que dependem do número de graus de liberdade do sistema. Tal comportamento permite uma explicação alternativa à dinâmica do processo de enovelamento, pois o mesmo sugere que o processo de enovelamento segue caminhos preferenciais ao longo da hiper-superfície de energia. Portanto, na busca de seu estado nativo a proteína não precisa visitar todos os estados acessíveis, mas somente aqueles associados ao espectro $f(\alpha)$.

Neste trabalho, estamos mostrando que, em função do tempo, as próprias trajetórias, seguidas pelas proteínas sobre essas regiões permitidas, são também multifractais. Neste cenário, a proteína, em sua evolução dinâmica, vai sendo influenciada pelo surgimento de estruturas intermediárias, que, gradativamente, por sucessivos aumentos de estabilidade conformacional, desviam as trajetórias pelos caminhos preferenciais de enovelamento. Consequentemente, a extrema facilidade com que uma proteína se enovela, apesar do número gigantesco de configurações possíveis, talvez possa ser atribuída à sucessão de eventos, experimentados pela mesma, numa hiper-superfície de energia, espaço-temporal, multifractal.

Apêndice A

Cálculo do expoente de escala ν para o modelo de Flory

Nosso objetivo neste apêndice é obter uma relação de escala entre o raio R e o tamanho do sistema N, utlizando a argumentação de campo médio apresentada por Flory [82], para um polímero embebido num bom solvente. Neste modelo, as forças envolvidas no empacotamento da estrutura são de duas naturezas: um entrópica (ou elástica) e outra de repulsão. Consideremos uma cadeia polimérica, composta por N monômeros e de raio R, mergulhada num espaço d-dimensional a energia desse sistema envolve essencialmente dois termos:

$$E = E_{ela} + E_{exc}, \tag{A.1}$$

onde $E_{ela} = \alpha(\frac{R}{R_0})^2$ descreve o comportamento plástico do sistema para grandes valores de R, $E_{exc} = \beta R^d \rho^2$ descreve o comportamento de exclusão do sistema, para

pequenos valores de R. Aqui ρ representa a densidade média do polímero,

$$\rho \sim \frac{N}{R^d}, \tag{A.2}$$

R_0 é o tamanho característico do sistema para grandes comprimentos ($R_0 \sim N^{1/2}$), enquanto que α e β são constantes numéricas que fornecem a dimensão apropriada de energia para a grandeza final. Assim:

$$E(R) = \alpha(\frac{R}{R_0})^2 + \beta R^d \rho^2, \tag{A.3}$$

ou ainda,

$$E(R) = \alpha\frac{R^2}{N} + \beta R^d \frac{N^2}{R^{2d}} = \alpha N^{-1} R^2 + \beta N^2 R^{-d}. \tag{A.4}$$

Desta forma a estrutura possui um raio característico de equilíbrio dado pela minimização da energia total,

$$\frac{\partial E}{\partial R} = 2\alpha N^{-1} R + -D N^2 R^{-(1+d)} = 0, \tag{A.5}$$

logo o comportamento de escala de R com N é dado por:

$$R \sim N^{\nu}, \tag{A.6}$$

onde

$$\nu = \frac{3}{d+2}. \tag{A.7}$$

Em particular, no caso em que d=3 temos $\nu = (3/5) = 0.60$.

Apêndice B

Determinação das coordenadas cartesianas do caminhante

Para determinarmos as coordenadas do resíduo subsequente, em termos da posição do resíduo anterior do caminhante discutido na Seção 3.2, utilizamos uma adaptação do método proposto por Levitt e colaboradores[87]. Na abordagem apresentada nesa Tese, a construção de versores ortogonais: **u**, **v** e **w**; determinam a rotação entre os planos formados pelo conjunto de três resíduos consecutivos. Assim, para um conjunto de ângulos de torção (ϕ_i,ψ_i) geramos as coordenadas dos carbonos C_α de 1 até N da seguinte forma:

1. As coordenadas dos **três** primeiros C_α são fixadas

$$\begin{aligned} \mathbf{r_1} &= (0.0, 0.0, 0.0); \\ \mathbf{r_2} &= (l_o, 0.0, 0.0); \\ \mathbf{r_3} &= (x_2 + l_o\cos(\pi - \phi_1), y_2 + l_o\mathrm{sen}(\pi - \phi_1), 0.0); \end{aligned} \tag{B.1}$$

aqui $l_o = 3.8$ representa o comprimento da ligação entre os C_α.

2. Em seguida as coordenadas de cada novo resíduo são calculadas a partir dos três precedentes. Para tanto calculamos, inicialmente, os vetores unitários : **u**, **v** e **w**

$$\mathbf{u} = \frac{\mathbf{r}_{i-2} - \mathbf{r}_{i-1}}{|\mathbf{r}_{i-2} - \mathbf{r}_{i-1}|}, \tag{B.2}$$

$$\mathbf{v} = \frac{(\mathbf{r}_{i-3} - \mathbf{r}_{i-2}) - [(\mathbf{r}_{i-3} - \mathbf{r}_{i-2})\mathbf{u}]\mathbf{u}}{|(\mathbf{r}_{i-3} - \mathbf{r}_{i-2}) - [(\mathbf{r}_{i-3} - \mathbf{r}_{i-2})\mathbf{u}]\mathbf{u}|}, \tag{B.3}$$

e

$$\mathbf{w} = \mathbf{u} \times \mathbf{v}. \tag{B.4}$$

O vetor **u** é o versor ao longo da pseudo-ligação entre C^α_{i-2} e C^α_{i-1}. Os vetores **u** e **v** definem juntos o plano contendo os átomos C^α_{i-3}, C^α_{i-2} and C^α_{i-1}. O vetor **w** completa a trinca do sistema de coordenadas e desta forma as coordenadas do próximo resíduo são dadas por:

$$\mathbf{r_i} = \mathbf{r_{i-1}} + l_o\cos(\pi - \phi)\mathbf{u} + l_o\mathrm{sen}(\pi - \phi)\cos(\psi)\mathbf{v} + l_o\mathrm{sen}(\pi - \phi)\mathrm{sen}(\psi)\mathbf{w}. \tag{B.5}$$

Bibliografia

[1] R. Mantegna and H. E. Stanley. *An Introduction to Econophysics: Correlations and Complextity in Finance*. Cambridge University Press, Cambridge, 1999.

[2] Edited by A. T. Skjeltorp and A. V. Belushkin. *Forces, Growth and Form in Soft Condensed Matter: At the Interface Between Physics and Biology*. Kluwer Academic Publishers, Norway, 2004.

[3] R. M. Zorzenon dos Santos and S. Coutinho. Dynamics of the HIV Infection: *a cellular automata*. *Physical Review Letters*, 87:168102, 2001.

[4] P. H. Figueirêdo, S. Coutinho, and R. M. Zorzenon dos Santos. *Robustness of a Cellular Automata Model for the HIV Infection*. Submetido para Physica A (2006). Código BA90: *status* ver http://www.if.ufrgs.br/barbosa/).

[5] Mauro Copelli, Antônio C. Roque, Rodrigo F. Oliveira, and Osame Kinouchi. Physics of psychophysics: Stevens and Weber-Fechner laws are transfer functions of excitable media. *Physical Review E*, 65:060901, 2003.

[6] Jr. J. S. Andrade, U. M. S. Costa, and M. L Lyra. Generalized Zipf's Law in Proportional Voting Processes. *Europhysics Letters*, 62:131–137, 2003.

[7] Viviane M. de Oliveira, M. A. F. Gomes, and I. R. Tsang. Theoretical Model for the evolution of the linguistic diversity. *Physica A*, 361:361–370, 2006.

[8] Fabiano Sobreira and Marcelo Gomes. The geometry of slums: boundaries, packing & diversity. *Centre for Advanced Spatial Analysis - Working Paper Series*, 30:131–137, 2001.

[9] Thomas Surrey, François Nédelec, Stanislas Leibler, and Eric Karsenti. Physical properties determining self-organiztion of motors and microtubules. *Science*, 292:1167–1171, 2001.

[10] G.F. Zebende, M.G. Pereira, E. Jr. Nogueira, and M.A. Moret. Universal persistence in astrophysical sources. *Physica A*, 349:452, 2005.

[11] Galileu Galilei. *Dialogues Concerning the two New Sciences, in Great Books of the Western World*. McGraw-Hill, Chicago, EUA, 1952.

[12] Isaac Newton. *Optica*. Edusp, São Paulo, Brasil, 2000.

[13] Erwin Schröndiger. *What is Life ?* Cambridge University Press, Cambridge, 1944.

[14] R. H. Pain (Ed.). *Frontiers in molecular biology: Mechanisms of protein folding - 2nd ed.* Oxford University Press, UK, 2000.

[15] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426:884–890, 2003.

[16] G. Taubes. Misfolding the way to disease. *Science*, 271:1493, 1996.

[17] W. Poon. Soft condensed matter: where physics meets biology. *PhysicsWeb*, 2001.

[18] Paul Davies. *O quinto milagre: em busca da origem da vida*. Companhia das Letras, São Paulo, Brasil, 2000.

[19] Michael Daune. *Molecular Biophysics: structures in motion*. Oxford University Press, New York, EUA, 1999.

[20] F. Richards. The protein folding problem. *Scientific American*, pages 34–41, 1991.

[21] P. H. Figueirêdo, M. A. Moret, E. Nogueira Jr., and S. Coutinho. *Gaussian distribution driving protein folding*. Submetido para Physical Review E: Rapid Communication(2006)(código LP10790ER).

[22] P. H. Figueirêdo, S. Coutinho, M. A. Moret, and E. Nogueira Jr. *Gaussian distribution driving protein folding: the role of stochasticity on compactness*. A ser submetido para Physical Review E (2006).

[23] P. H. Figueirêdo, E. Nogueira Jr., M. A. Moret, and S. Coutinho. *Multifractal Analysis of Protein Energy Time Series*. A ser submetido para Physical Review E (2006).

[24] Abdus Salam. The role of Chirality in the origin of life - Chirality, Phase Transitions and their Introduction in Aminoacids. *International Centre for Theoretical Physics - College on Methods and Experimental Techniques in Biophysics*, 1992.

[25] Henrik Flyvbjerg, John Hertz, Mogens H. Jensen, Oleg G. Mouritsn, and Kim Sneppen (Eds.). *Lecture Notes in Physics - Physics of Biological systems: from molecules to species*. Springer, Germany, 1997.

[26] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29:7133, 1990.

[27] Paulo M. Bisch and Pedro G. Pascutti. *Métodos Computacionais em Biologia - Modelagem de Biomoléculas*. II Escola de verão - IBCCF - UFRJ, Petrópolis, 2000.

[28] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. *J. Mol. Biol.*, 24:95, 1963.

[29] G. N. Ramachandran. Molecular forces in protein structure and crystallography. *Int. J. Protein Res.*, 1(1):5–17, 1969.

[30] G. N. Ramachandran and R. Chandrasekharan. The conformation energy map of a dipeptide unit in relation to infra red and nuclear magnetic resonance data. *Biopolymers*, 10:935, 1971.

[31] Ricardo Ferreira. *Watson & Crick: a história da descoberta da estrutura do DNA*. Odyseus Editoras Ltda., Brasil, 2003.

[32] P. L. Privalov and G. I. Makhatadze. Contridution of hydration and noncovalent interactions to the heat capacity effect on protein unfolding. *J. Mol. Biol.*, 224:715–723, 1992.

[33] Audum Bakk and Johan S. Hϕye. Microscopic argument for the anomaulous hydration heat capacity increment upon solvation of polar substances. *Physica A*, 303:286–294, 2002.

[34] Audum Bakk, Johan S. Hϕye, and Alex Hansen. Specific heat upon aqueous unfolding of the protein interior: a theoretical approach. *Physica A*, 304:355–361, 2002.

[35] D. L. Stein. *Decoherence and Entropy in Complex Systems, Edited by H. T. Elze, Lecture Notes in Physics*, volume 633. Springer, 2003.

[36] D. L. Stein. *Spin Glasses and Biology*. Word Scientific, Singapore, 1992.

[37] H. Wu. *Am. J. Physiol.*, 90(562), 1929.

[38] H. Wu. *Chinese J. Physiol.*, 5(321), 1931.

[39] A. E. Mirsky and L. Pauling. Structure of native, denatured and coagulated proteins. *Proc. Natl. Acad. Sci.*, 22:439–447, 1936.

[40] C. Anfisen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[41] J. L. Shol, S. S. Jaswal, and D. A. Agard. Unfolded conformations of α-lytic protease are more stable than its native sate. *Nature*, 395:817, 1998.

[42] Vladmir N. Uversky. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*, 11:739–756, 2002.

[43] C. J. Levinthal. Are there pathways for protein folding ? *Journal of Chemical Physics*, 65:44–45, 1968.

[44] C. Levinthal. *In Mösbauer Spectroscopy in biology systems - How to fold graciously*. Monticello, Illinois, 1969.

[45] M. Karplus. The Levinthal Paradox: yesterday and today. *Fold. Des*, 2:S69, 1997.

[46] Robert Zwanzig, Attila Szabo, and Biman Baggachi. Levinthal´s paradox. *Proc. Natl. Acad. Sci.*, 89:20–22, 1991.

[47] E. I. Shakhonovich. Theoretical studies of protein-folding thermodynamics and kinetcs. *Curr. Opin. Strut. Biol.*, 7:29–40, 1997.

[48] Carl Branden & John Tooze. *Introduction to Protein Structure - 2nd Edition - pages 89-94*. Garland Publishing, New York, 1998.

[49] M. A. Moret, P. G. Pascutti, K. C. Mundim, P. M. Bisch, and E. Nogueira Jr. Multifractality, Levinthal paradox, and energy hypersurface. *Physical Review E*, 63:020901–020904, 2001.

[50] C.E. Bugg. Protein Crystallography. *International Centre for Theoretical Physics - College on Methods and Experimental Techniques in Biophysics*, 1992.

[51] K. Wuthrich. Protein Structure Dertemination in Solution by Nuclear Resonance Spectroscopy. *International Centre for Theoretical Physics - College on Methods and Experimental Techniques in Biophysics*, 1992.

[52] N. Sreerama, S.Y. Venyaminov, and R.W. Woody. Estimationof the number of alpha-helical and beta-strand segments in proteins using dichroism spestroscopy. *Protein Science*, 8, 1999.

[53] H. A. Kramers. Bronwnian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284, 1940.

[54] H. Risken. *The Fokker-Planck Equation*. Springer, 1996.

[55] P. Hänggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Review on Modern Pyhisics*, 62:251, 1990.

[56] R. Guerois, J. E. Nielsen, and L. Serrano. Predictingchanges in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, 320:369, 2002.

[57] N. Go. Studies on protein folding, unfolding and fluctuations by computer simulation. the effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.*, 7:445–459, 1975.

[58] J. Shimada, E. Kussel, and E. I. Shakhonovich. The folding thermodynamics and kinetics of cramin using an all-atom Monte Carlo simulation. *J. Mol. Biol.*, 308:79, 2001.

[59] S. Miyazawa and R. Jernigan. Estimation of effective interresidue contact energies from protein crystal structures. *Macromolecules*, 18:534, 1985.

[60] B. Derrida. Random energy model, an exactly solvable model of disordered systems. *Physical Review B*, 24:2613, 1981.

[61] E. I. Shakhonovich and A. Gutin. Enumaration of all compact conformations of copolymers with random sequence of links. *Journal of Chemical Physics*, 93:5967, 1989.

[62] J. D. Brygelson and P. G. Wolynes. Intermedites and barrier crossing in a random energy model. *Journal of Chemical Physics*, 93:6902, 1989.

[63] B. R. Brooks, R. E. Bruccoreli, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARRM: a program for macromolecular energy, mnimization, and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[64] S. J. Weiner, P. Kollman, D. T. Nguyen, and D. Case. An all atom force field for simulationsof proteins and nucleic acids. *J. Comp. Chem.*, 7:230–252, 1986.

[65] W. F. van Gunsteren and H. J. C. Berendsen. Groningen Molecular Simulation (GROMOS) - Library Manual, Biomos, Groningen. 1987.

[66] M. Clark, R. D. Cramer, and N. van Opdenbosch. Validation of the general purpose TRIPOS 5.2 force field. *J. Comp. Chem.*, 10:982–1012, 1989.

[67] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, and D. T. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and ernetic properties of molecules. 91:1, 1990.

[68] S. S.Petrova and A. D. Solov'ev. The origin of the method of steepest descent. *Historia Mathematica*, 24(4):361–375, 1997.

[69] W. H. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *"Minimization or maximization of functions, " in Numerical Recipes in C.* Cambridge University Press, Cambridge, UK, 2000.

[70] M.A. Moret, P.G. Pascutti, P.M. Bisch, and K.C. Mundim. Stochastic Molecular Optimization Using Generalized Annealing. *J. Comp. Chem.*, 19(6):647–657, 1998.

[71] José Nelson Onuchic and Peter G. Wolynes. Theory of protein folding. *Curr. Opin. Strut. Biol.*, 14:70–75, 2004.

[72] Kit Fun Lau and Ken A. Dill. A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of proteins. *Macromolecules*, 22:3986–3997, 1989.

[73] C. Tang. Simple models of the protein folding problem. *Physica A*, 288:31–48, 2000.

[74] R. A. Broglia and G. Tiana. Hierarchy of events in the folding of model proteins. *Journal of Chemical Physics*, 114:7267, 2001.

[75] R. Brown. A brief account of microscopical observation made in the months of june, july and august, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies.). *Philosophical Magazine*, 4:161, 1828.

[76] Louis Bachelier. Théorie de la Spéculation. *Annales Scientifiques de l´Ecole Normale Supérieure*, 17:21, 1900.

[77] A. Einstein. Uber die von der molekularkinetischen theorie der wärme gefordete bewegung von in ruhenden flüssigkeiten suspendierten teilchen (On the motion required by the molecular kinetic theory of heat of small particles suspended in stationary liquid). *Annalen der Physik*, 17:549, 1905.

[78] D. Chowdhury and B. K. Chakrabarti. Random walk on self-avoiding walk: a model for conductivity of linear polymers. *Journal Physics A: Math. Gen.*, 18:L377–L382, 1985.

[79] R. H. A. David Shaw and J. A. Tuszynski. Random walks with non-Gaussian step-size distributions and the folding of random polymer chains. *Physical Review E*, 67:031102–1–031102–11, 2003.

[80] J. D. Noh H. ParK D. Kim and M. den Nijs. Anomalous roughness, localization, and globally constrained random walks. *Physical Review E*, 64:046131.1–046131.14, 2001.

[81] C. C. Donato, M. A. F. Gomes, and F. A. OLIVEIRA. Anomalous diffusion on crumpled wires in two dimensions. *Physica A*, 368, 2006.

[82] P. J. Flory. The configuration of real polymer chains. *Journal of Chemical Physics*, 17:303, 1949.

[83] Kerson Huang. *Lectures on Statistical Physics and Protein Folding.* World Scientific, Singapore, 2005.

[84] B. Li, N. Madras, and A. D. Sokal. Critical exponents, hyperscaling and universal amplitude ratios for two-and three-dimensional self-avoiding walks. *J. Stat. Phys.*, 80:661, 1995.

[85] M. A. F. Gomes. Fractal geometry in crumpled paper balls. *Americam Journal Physics*, 55(7):649–650, 1987.

[86] M. J. Rooman, J. P. Kocher, and S. J. Wodak. Prediction of protein backbone conformation based on seven structure assignments : Influence of local interactions. *J. Mol. Biol.*, 221:961, 1991.

[87] Britt H. Park and Michael Levitt. The Complexity and Accuracy of Discrete State Models of Protein Structure. *J. Mol. Biol.*, 249:493–507, 1995.

[88] Barry Honig. Protein Folding: From the Levinthal Pradox to Structure Prediction. *J. Mol. Biol.*, 292:283–293, 1999.

[89] Igor N. Berezovsky and Edward N. Trifonov. Loop fold nature of globular proteins. *Protein Engineering*, 14:403–407, 2001.

[90] J. Zhang, R. Chen, C. Tang, and J. Liang. Origin of scaling behavior of protein packing density: A sequential Monte Carlo study of compact long chain polymers. *Journal of Chemical Physics*, 118(13):6102–6109, 2003.

[91] Amos Maritan, Cristian Micheletti, Antonio Trovato, and Jayanth R. Banavar. Optimal shapes of compact strings. *Nature*, 406:287–290, 2000.

[92] Amos Maritan and Jayanth R. Banavar. Colloquium: Geometrical approach to protein folding: a tube picture. *Review on Modern Pyhisics*, 75:23, 2003.

[93] Francesco Valle, Mélanie Favre, Paolo De Los Rios, Angelo Rosa, and Giovanni Dietler. Scaling exponents and probability distributionsof dna end-to-end distance. *Physical Review Letters*, 95:158105, 2005.

[94] M. A. Moret, P. M. Bisch, and F. M. C. Vieira. Algorithm for multiple minima search. *Physical Review E*, 57:R2535, 1998.

[95] J. Feder. *Fractals*. Plenum Press, New York, EUA, 1988.

[96] P. G. Pascutti, K. C. Mundim, A. S. Ito, and P. M. Bisch. Polarization effect on peptide conformations at water-membrane interface by molecular dynamics simulation. *J. Comp. Chem.*, 20:971, 1999.

[97] K. C. Mundim, P. G. Pascutti, and P. M. Bisch. *THOR - Force Field and Molecular Dynamics Simulations*. http://www.unb.br/iq/kleber/Thor/node1.html.

[98] D. A. Lidar, D. Thirumalai, R. Elbre, and R. B. Gerber. Fractal Analysis of Protein Energy Landscapes. *Physical Review E*, 59:2231, 2001.

[99] Marcelo A. Moret. *Estudo Conformacional de Proteínas Usando Métodos Estocásticos Generalizados.* Tese de Doutorado, Universidade Federal do Rio de Janeiro, Brasil, 2000.

[100] M. A. Moret, P. M. Bisch, E. Nogueira Jr., and P. G. Pascutti. Stochastic strategy to analyze protein folding. *Physica A*, 353:353, 2005.

[101] Klauss Schulten, Hui Lu, and Linsen Bai. *Probing Protein Motion Trhough Temperature Echoes in Lecture Notes in Physics - Physics of Biological systems: from molecules to species.* Springer, Germany, 1997.

[102] K. R. Shoemaker, P. S. Kim, E. J. Stewart, and R. L. Baldwin. Tests of the helix dipole model for stabilization of α-helices. *Nature*, 326:563–567, 1987.

[103] M.A. Moret, P.M. Bisch, K.C. Mundim, and P.G. Pascutti. New Stochastic Strategy to Analyze Helix Folding. *Biophysycal Journal*, 82:1123–1132, 2002.

[104] N.C. Rogers. *The role of electrostatic interactions in the structure of the globular proteins.* In *Prediction of Protein Structure and Principles of Protein Conformations.* . Plenum Press, New York, 1989.

[105] P. S. Kim and R. L. Baldwin. A kelix stop signal in the isolated S-peptide of ribonuclease A. *Nature*, 307:329–334, 1984.

[106] S. Marqusee, V. H. Robins, and R. L. Baldewin. Unusually stable helix formation in short alanine based peptides. *Proc. Natl. Acad. Sci.*, 86:5286–5290, 1989.

[107] P. Doruker and I. Bahar. Role of water on unfolding kinetics of helical peptides studied by molecular dynamics simulations. *Biophysycal Journal*, 72:2445–2456, 1997.

[108] A. Hitpold, P. Ferrara, J. Gsponer, and A. Caflisch. Free energy surface of the helical peptide Y(MEARA)(6). *Journal of Chemical Physics B.*, 104:10080–10086, 2000.

[109] T. Vicsek. *Fractal Growth Phenomena, 2nd. Ed.* World Scientific, Singapore, 1992.

[110] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267:1619, 1995.

[111] C.V. Chianca, A. Ticora, and T. J. P. Penna. Fourier-detrended fluctuation analysis. *Physica A*, 357:447, 2005.

[112] S. Havlin, S. V. Buldyrev, A. Bunde, A. L. Goldberger, P. Ch. Ivanov, C.-K. Peng, and H. E. Stanley. Scaling in nature: from dna through heartbeats to weather. *Physica A*, 316:46, 1999.

[113] A. Arneodo, E. Bacry, P.V. Graves, and J.F. Muzy. Characterizing Long-Range Correlations in DNA Sequences from Wavelet Analysis. *Physical Review Letters*, 74:3293, 1996.

[114] P. Manimaran, Prasanta K. Panigrahi, and Jitendra C. Parikh. Wavelet analysis and scaling properties of time series. *Physical Review E*, 72:046120, 2005.

[115] C. K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H.E. Stanley. Long-range correlations in nucleotide sequences. *Nature*, 356:168, 1992.

[116] P. Ch. Ivanov, L.A. Nunes Amaral, A.L. Goldberger, S. Havlin, M.G. Roseblum, Z.R. Struzik, and H.E. Stanley. Multifractality in human heartbeat dynamics. *Nature*, 399:461, 1999.

[117] E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Roman, Y. Goldreich, and Schellnhuber. Indication of a universal persistence law governing atmospheric variability. *Physical Review Letters*, 81:729, 1998.

[118] M.A. Moret, G.F. Zebende, E.Jr. Nogueira, and M.G. Pereira. Fluctuations analysis of stellar x-ray binary systems. *Physical Review E*, 68:041104, 2003.

[119] Pawel Oswiecimka, Jaroslaw Kwapien, and Stanislaw Drozdz. Wavelet versus detrended fluctuation analysis of multifractal structures. *Physical Review E*, 74:016103, 2006.

[120] J.W. Kanterlhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H.E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316:87, 2002.

[121] Alexandre Rosas, Edvaldo Nogueira Jr., and José F. Fontanari. Multifractal Analysis of DNA walks and trails. *Physical Review E*, 66:061906, 2002.

[122] R.G. Kavasseri and R. Hagarajan. A multifractal description of wind speed records. *Chaos, Solitons and Fractals*, 24:165–173, 2005.

[123] L. Telesca, V. Lapenna, and M. Macchiato. Self-similarity properties of seismicity in the southern aegean area. *Tectonophysics*, 321:179–188, 2000.

[124] J. D. Brygelson and P. G. Wolynes. A simple statistical field-theory of heteropolymer collapse with application to protein folding. *Biopolymers*, 30(015102):177, 1990.

[125] M. Prévost and I. Ortmans. Refolding simulation of an isolated fragment of barnase into a native-like β-harpin: evidence for compactness and hydrogen bonding as concurrent stabilizing factors. *Proteins*, 29:212, 1997.