

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO  
DEPARTAMENTO DE FÍSICA E MATEMÁTICA  
PÓS-GRADUAÇÃO EM FÍSICA APLICADA À MEDICINA E BIOLOGIA

## **Análise de modos normais em proteínas**

Matheus Rodrigues de Mendonça

Ribeirão Preto - SP  
2010

MATHEUS RODRIGUES DE MENDONÇA

## **Análise de modos normais em proteínas**

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para a obtenção do título de Mestre em Ciências

**Área de concentração:**

Física Aplicada à Medicina e Biologia

**Orientador:**

Prof. Dr. Nelson Augusto Alves

Ribeirão Preto - SP

2010

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

MENDONCA, M. R.

Análise de modos normais em proteínas / Matheus Rodrigues de Mendonça; orientador Nelson Augusto Alves – Ribeirão Preto/SP, 2010.

71 p.

Dissertação (Mestrado – Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia) – Departamento de Física e Matemática da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo.

1. Introdução. 2. Modelos de rede elástica. 3. Métodos. Resultados e Discussão. 4. Conclusões e Perspectivas.

*Dedico esta dissertação à minha amada família,  
em especial, ao meu sobrinho Lucas de um ano.*

# Agradecimentos

Primeiramente, agradeço a Deus, pois sem Ele nada seria possível.

Aos meus pais, Sônia Maria Rodrigues de Mendonça e Mailton Ribeiro de Mendonça, por me proporcionarem uma vida com dignidade. E ao meu irmão, Adriano Rodrigues de Mendonça, por me motivar a estudar.

Ao Professor Doutor Nelson Augusto Alves, pela sua excelente orientação, amizade, pelo exemplo de profissionalismo e, principalmente, pelo rico conhecimento e aprendizado que adquiri, por meio dele, durante o mestrado.

Ao professor Doutor Alexandre Martinez, pela amizade e supervisão no estágio PAE.

Aos amigos do grupo de pesquisa, em especial ao Leandro G. Rizzi, pela amizade, pelo apoio técnico e pelas instruções sobre Latex e Xmgrace.

Ao Rafael Frigori, pela amizade e pelos comentários pertinentes nos seminários apresentados ao grupo de pesquisa.

A Jacyana pela amizade, carinho e pelo bolo de chocolate no meu aniversário.

Ao Guidolin, pela amizade e pelo auxílio técnico na instalação de bibliotecas e na manutenção do cluster de computadores.

Às secretárias do Departamento de Física e Matemática, Nilza e Sônia, pelos serviços administrativos realizados com competência e educação.

Ao Professor Doutor Marcelo Mulato, atual coordenador do programa FAMB, pela sua excelente gestão e por viabilizar auxílio financeiro para participação de congressos.

Ao Professor Doutor Vitor Barbanti Pereira Leite e à professora Doutora Silvana Giuliatti, por terem aceitado o convite para participar da minha banca de Mestrado e, sobretudo, pelas contribuições para este trabalho.

A Juliana Alencar, minha querida amiga, pelo carinho e por sempre estar ao meu lado com prontidão para me ajudar naquilo que estava ao seu alcance, especialmente, pela leitura do texto.

À minha querida amiga Hurzana Mello e ao seu irmão Matheus Mello por assistirem à minha defesa de mestrado.

A Elizabete Rosa e Nathalia dos Reis pelas conversas agradáveis e motivadoras durante as caminhadas no campus da USP.

Aos demais colegas e amigos que fiz no decorrer do mestrado:

- do Programa de Pós-graduação em Física Aplicada à Medicina e Biologia (FAMB): Ao Olavo Menin, Marcelo Pereira “Marcelão”, Brenno Cabella, Juliana Berbert, Tiago Arruda “Turco”, Tobias Heimfarth, Aquino Espindola, Raimundo Nonato, Ariadne Costa, Diogo Vieira, Fernanda Oliveira, Lindomar dos Santos e Natalia Destefano;
- da Aliança Bíblica Universitária (ABU): Sara Tironi, Eder Coradi, Jonatas Jané, Jonatas Bloch, Renata Ferrari, Pedro Paulo, Milena Naomi, Rafael Santos, David Tatsuo “Ouriço” e Marina Batistuti “Net”.

Os quais contribuíram em tornar o ambiente de estudo e trabalho mais agradável.

À agência de fomento CAPES, pelo apoio financeiro.

*“ Física é uma ciência de modelos e a modelagem é uma atividade sistemática dos físicos para construir e aplicar o conhecimento científico. Aprender Física, implica, então, aprender a jogar o jogo da modelagem”.*  
**(MOREIRA, 1997)**

# Resumo

A abordagem de modos normais de baixa frequência na descrição das flutuações conformacionais dos estados nativos das proteínas globulares tem ajudado na caracterização das suas funções biológicas. Vários métodos teóricos e experimentais têm sido empregados para a determinação destas flutuações internas. Estes movimentos podem ser caracterizados pelo fator Debye-Waller (fator-B), correspondente à mobilidade local do resíduo em nível atômico. A análise de modos normais utilizando os modelos de rede elástica (ENM) demonstra ser uma técnica robusta. Fatores-B experimentais são reproduzidos teoricamente por meio desta técnica em tempos computacionais relativamente curtos, mostrando-se competitiva com as técnicas mais sofisticadas. O modelo de rede elástica é uma abordagem tipo *coarse-grain* na qual a proteína no seu estado enovelado é representada por uma rede elástica tridimensional de carbonos  $\alpha$  ( $C_\alpha$ ) conectados por molas. As molas representam as interações ligantes e não ligantes entre os  $C_\alpha$ . Neste trabalho, inicialmente, estudamos os modelos de rede elástica já conhecidos na literatura. Em seguida, realizamos um estudo comparativo entre eles. Neste estudo, comprovamos que os modelos pfGNM e pfANM apresentam melhor correlação com os fatores-B experimentais que os os modelos GNM e ANM tradicionais. Desenvolvemos também uma nova abordagem, a qual intitulamos número de contatos ponderados anisotrópica (AWCN). Mostramos que a abordagem AWCN apresenta um desempenho significativamente melhor que o modelo de rede elástica anisotrópica tradicional. Por fim, realizamos um estudo de caráter investigativo do comportamento do peso das interações entre resíduos. Este estudo revelou que, para os modelos WCN e AWCN, a correlação exibe o seu valor máximo para interações ponderadas  $R_{ij}^{-p}$ , entre resíduos  $i$  e  $j$ , para valores de  $p$  em torno de 2. Nos modelos pfGNM e pfANM a correlação é maximizada para dois valores de  $p$ , o primeiro em torno de 2 e o segundo em torno de 4,75, indicando que a ponderação pelo recíproco do quadrado da distância, usualmente empregada na literatura, pode não ser adequada para obter a melhor correlação.



# Abstract

Low frequency normal mode approach to describe conformational fluctuations of globular proteins has helped to characterize their biological functions. Various theoretical and experimental methods have been employed to determine the magnitudes of those internal motions. Those motions can be characterized by the Debye-Waller factor (B-factor), corresponding to the local mobility of the residue at the atomic level. Normal mode analysis using elastic network models (ENM) has demonstrated to be a robust technique. Experimental B-factors has been reproduced theoretically by means of this technique in a short computational time and it has been shown to be competitive with more sophisticated techniques. The ENM is a coarse-grained approach in which the protein is represented by a three-dimensional elastic network of alpha-carbon atoms connected by springs. Springs represent bonded and non-bonded interactions between the alpha-carbon atoms. In this work, we study the elastic network models known in the literature. Next, we perform a comparative study between them. We show that the pfGNM and pfANM models present better correlation with experimental B-factors than the traditional GNM and ANM models. We also develop a new approach, which we entitled anisotropic weighted contact number (AWCN). We show that it presents results significantly better than the traditional anisotropic elastic network model. Finally, we perform a study of investigative character of the behavior for the weight of the interactions between residues. This study revealed that, for the WCN and AWCN models, the correlation exhibits its maximum value for weighted interactions  $R_{ij}^{-p}$ , between residues  $i$  and  $j$ , for values of  $p$  around 2. In the pfGNM and pfANM models the correlation is maximized for two values of  $p$ , the first one around 2 and the second one around 4.75. This indicates that the weighting by the reciprocal of the square of the distance, usually employed in the literature, may not be appropriate to obtain the best correlation.

# Sumário

<b>Lista de Figuras</b>	<b>11</b>
<b>Lista de Tabelas</b>	<b>13</b>
<b>1 Introdução</b>	<b>14</b>
<b>2 Modelos de rede elástica (ENM)</b>	<b>19</b>
2.1 Modelo proposto por Gō . . . . .	19
2.2 Modelo proposto por Tirion . . . . .	21
2.3 Modelo de rede gaussiana (GNM) . . . . .	21
2.4 Modelo de rede gaussiana (ANM) . . . . .	28
2.4.1 Comparando fatores-B anisotrópicos com dados experimentais . . . . .	29
2.5 Número de contatos ponderados (WCN) . . . . .	30
2.6 Modelos de rede elástica livres de parâmetro (pfGNM e pfANM) . . . . .	32
<b>3 Métodos</b>	<b>33</b>
3.1 Banco de dados de proteínas . . . . .	33
3.2 Programas computacionais de análise de modos normais . . . . .	34
<b>4 Análise comparativa entre os modelos de rede elástica</b>	<b>36</b>
<b>5 Novos modelos</b>	<b>41</b>
5.1 Modelo AWCN . . . . .	41
5.2 Expoente generalizado no potencial de interação entre resíduos . . . . .	42

<b>6</b>	<b>Conclusões e perspectivas</b>	<b>44</b>
	<b>Referências</b>	<b>46</b>
	<b>Apêndice A – Análise de modos normais (NMA)</b>	<b>48</b>
	A.1 Teoria clássica dos modos normais . . . . .	48
	<b>Apêndice B – Demonstrações</b>	<b>51</b>
	B.1 Demonstração do cálculo de $H_{ij}$ . . . . .	51
	B.2 Condição de normalização . . . . .	52
	<b>Apêndice C – Teoria de grafos</b>	<b>54</b>
	C.1 Uma breve introdução . . . . .	54
	C.1.1 A matriz Laplaciana . . . . .	54
	C.2 Matriz de Kirchhoff . . . . .	56
	C.3 Cálculo dos elementos da matriz Hessiana no ANM . . . . .	56
	<b>Apêndice D – Programas</b>	<b>59</b>
	D.1 GNM . . . . .	59
	D.2 pfGNM . . . . .	66
	<b>Apêndice E – Lista das proteínas analisadas</b>	<b>68</b>

# Lista de Figuras

1	Representação diagramática da estrutura tridimensional da proteína cujo código PDB é 1PGA. A hélice $\alpha$ está ilustrada em vermelho, as fitas $\beta$ em amarelo e os contornos na cor verde. Esta figura foi obtida por meio do aplicativo PyMOL. . . . .	15
2	(a) Representação da cadeia linear de resíduos ( <i>backbone</i> ). (b) representação do <i>backbone</i> ilustrada pela linha mais escura e da rede elástica tridimensional de resíduos pela linha mais clara. Figura adaptada do artigo (ATILGAN e colaboradores., 2001). . . . .	20
3	(a) Representação da proteína por uma rede elástica tridimensional, onde os vértices da rede são definidos pelas coordenadas dos resíduos e as molas representam as interações não ligantes entre eles. Neste modelo, considera-se que todas as molas possuem constante de força $\gamma$ . (b) representação do sistema de referência fixo no laboratório para descrever as posições dos resíduos $i$ e $j$ . . . . .	22
4	Energia potencial de um tripeptídeo. Como a distância entre o resíduo 1 e o 3 é maior que $R_c$ , desconsidera-se esta interação. A equação final da Hamiltoniana $\mathcal{H}_{\text{GNM}}$ na forma matricial é semelhante a equação (2.7). . . .	24
5	Comparação entre os espectros do fator de temperatura teórico determinado pelo GNM e o experimental. . . . .	27
6	Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo GNM para proteínas das classes <i>all-<math>\alpha</math></i> , <i>all-<math>\beta</math></i> , $\alpha + \beta$ e $\alpha/\beta$ . . . . .	37
7	Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo ANM para proteínas das classes <i>all-<math>\alpha</math></i> , <i>all-<math>\beta</math></i> , $\alpha + \beta$ e $\alpha/\beta$ . . . . .	38

8	Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo pfANM para proteínas das <i>all-<math>\alpha</math></i> , <i>all-<math>\beta</math></i> , $\alpha + \beta$ e $\alpha/\beta$ . . . . .	38
9	Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo pfGNM para proteínas das classes <i>all-<math>\alpha</math></i> , <i>all-<math>\beta</math></i> , $\alpha + \beta$ e $\alpha/\beta$ . . . . .	39
10	Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo WCN para proteínas das classes <i>all-<math>\alpha</math></i> , <i>all-<math>\beta</math></i> , $\alpha + \beta$ e $\alpha/\beta$ . . . . .	39
11	Comportamento do coeficiente de correlação $c$ em função do parâmetro $p$ , determinado pelos modelos WCN, pfGNM, pfANM e AWCN. Estas curvas estão associadas à proteína cujo código PDB é 16PK, a qual contém 425 resíduos. . . .	43

# Lista de Tabelas

1	Coefficiente de correlação médio $\bar{c}$ calculado pelos modelos GNM, ANM, pfANM, pfGNM e WCN para as estruturas distribuídas de acordo com a classificação SCOP. O valor do raio de corte é representado no subscrito de $\bar{c}$ para os modelos que utilizam este parâmetro. . . . .	36
2	Proporção das proteínas que obtiveram valores de $c$ maiores que 0,5. . . . .	40
3	Coefficiente de correlação médio para cada modelo apresentado e para as estruturas distribuídas de acordo com a classificação SCOP. . . . .	42
4	Código PDB e o número de resíduos $N$ para 125 proteínas em cada classe de acordo com a classificação SCOP. . . . .	68

# 1 Introdução

Proteínas correspondem à classe mais comum de macromoléculas e são de vital importância em processos bioquímicos nos seres vivos. Formadas por longas cadeias lineares compostas pelos 20 aminoácidos que ocorrem naturalmente, as proteínas podem diferir em tamanho e em termos dos aminoácidos que as constituem. A especificação da sequência de aminoácidos é chamada de estrutura primária da proteína. As cadeias de aminoácidos formam, localmente, estruturas regulares tais como hélices  $\alpha$ , folhas  $\beta$  e contornos. Estas estruturas regulares são conhecidas como estruturas secundárias da proteína. A figura 1 ilustra a estrutura tridimensional da proteína cujo código PDB é 1PGA, a qual apresenta quatro fitas  $\beta$ , uma hélice  $\alpha$  e contornos.

É bem conhecido que as proteínas globulares evoluem de forma espontânea para conformações compactas únicas, cujo mecanismo chamado de enovelamento de proteínas (*protein folding*) é ainda pouco compreendido. Inúmeros estudos, teóricos e experimentais (SHAKHNOVICH, 1997; HENZLER-WILDMAN; KERN, 2007; MCCAMMON; GELIN; KARPLUS, 1997), continuam a ser realizados com a finalidade de identificar os possíveis mecanismos envolvidos. As estruturas compactas são chamadas de terciárias e correspondem ao estado biologicamente ativo. Tudo indica que este enovelamento espontâneo, conduzindo à formação da estrutura terciária, é determinado unicamente pela sua sequência primária (ORENGO; THORNTON, 2005). As interações entre as cadeias laterais dos aminoácidos são importantes para a estabilidade das proteínas e determinação da função que desempenham em processos biológicos, além de estabelecer a dinâmica do processo de enovelamento. Essas interações, não covalentes entre cadeias laterais, fazem com que haja a estabilização da orientação mútua das estruturas secundárias no estado completamente enovelado, correspondendo ao estado nativo da proteína.

Nossa proposta utiliza como ponto de partida uma representação de rede para as proteínas (GREENE; HIGMAN, 2003). Esta abordagem considera que uma proteína é constituída somente pelos carbonos  $\alpha$  ( $C_\alpha$ ), os quais passam a ser os vértices de uma rede e, as bordas representam as interações covalentes e não-covalentes com alcance até determinada

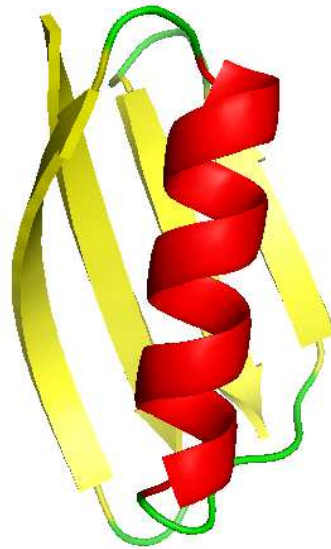


Figura 1: Representação diagramática da estrutura tridimensional da proteína cujo código PDB é 1PGA. A hélice  $\alpha$  está ilustrada em vermelho, as fitas  $\beta$  em amarelo e os contornos na cor verde. Esta figura foi obtida por meio do aplicativo PyMOL.

distância. Outra possibilidade é considerar o centro de massa das cadeias laterais como sendo os vértices da rede (ALVES; MARTINEZ, 2007), conforme estudo desenvolvido pelo nosso grupo. Tendo a descrição de rede, utilizaremos a abordagem de modos normais para extrair informação sobre movimentos funcionais de “larga escala” como o fator-B associado ao estado conformacional nativo (HALILOGLU; BAHAR; ERMAN, 1997).

A flexibilidade das proteínas no estado nativo está associada a diversos processos biológicos como atividade catalítica e reconhecimento biológico. Na seqüência protéica existem regiões flexíveis que podem ser medidas. Estas regiões dinâmicas são medidas pelo fator Debye-Waller (fator-B), correspondente à mobilidade local do resíduo em nível atômico. Este fator representa a diminuição da intensidade na difração por causa da desordem dinâmica da molécula, resultante da vibração dos átomos dependentes da temperatura. Tanto maior será a desordem quanto maior for a temperatura da proteína.

O estudo da dinâmica das proteínas é muito importante por apresentar informações relevantes sobre o seu mecanismo funcional. Por meio de técnicas teóricas e experimentais como dinâmica molecular, cristalografia de raios-x e ressonância magnética de neutrons, se obtêm informações sobre a estrutura e dinâmica da proteína (TIRION, 1996). Entretanto, essas técnicas demandam alto custo computacional e/ou experimental, além de serem lentas.

A análise de modos normais (NMA) de moléculas poliatômicas é uma técnica tradi-



onal em espectroscopia molecular. Um método bastante utilizado pelos espectroscopistas é o método da matriz GF (WILSON; DECIUS; CROSS, 1955). Neste método, as interações entre os átomos são consideradas harmônicas e as vibrações são descritas por coordenadas internas. Além disso, simplificações são feitas por meio de argumentos de simetria da teoria de grupo. Detalhes do sistema, como a geometria da molécula, os ângulos entre as ligações e a natureza das ligações são relevantes neste método. Desde a década de 50, a NMA tem sido efetuada pelo método da matriz GF em moléculas pequenas e tem demonstrado ser uma ferramenta poderosa na predição de espectros de frequências vibracionais.

Os primeiros estudos da NMA em proteínas foram reportados na década de 80. Inicialmente, estes estudos foram realizados por meio de simulações de dinâmica molecular utilizando campos de forças altamente complexos e tendo como variáveis dinâmicas as coordenadas internas. Os modos normais e as frequências normais de vibração eram calculados via diagonalização da matriz Hessiana (matriz das constantes de força). Pelo fato de que os elementos da matriz Hessiana são calculados na conformação de equilíbrio, era necessário efetuar a minimização da energia.

Gō e colaboradores (GŌ; NOGUTI; NISHIKAWA, 1993) estudaram a dinâmica da proteína bovina pancreática tripsina inibidora (BPTI) utilizando como variáveis dinâmicas apenas os ângulos torcionais entre os planos de ligações e potenciais relativamente mais simples do que os usuais. Eles consideraram os ângulos torcionais como sendo as únicas variáveis dinâmicas. Isto decorre do fato de que as variações dos comprimentos e dos ângulos das ligações serem muito menores que as variações associadas aos ângulos torcionais. Assim, as quantidades comprimento e ângulos de ligações foram consideradas fixas. A vantagem de diminuir o número de variáveis dinâmicas está na redução do tamanho da matriz Hessiana e, conseqüentemente, na obtenção de maior ganho de eficiência computacional.

Computacionalmente, um grande empecilho nos estudos relatados anteriormente encontrava-se na etapa da minimização da energia. Assumindo a conformação nativa como sendo a conformação de equilíbrio, a autora Tirion (TIRION, 1996) desenvolveu um modelo no qual esta etapa é eliminada. Além disso, nesse modelo as interações entre os átomos são descritas por um potencial harmônico com uma única constante de força, para todos os pares de átomos, escolhida por ajuste com o potencial experimental.

Inspirados no modelo desenvolvido por Tirion e na teoria de rede elástica de polímeros desenvolvida por Flory (FLORY, 1976), Haliloglu e colaboradores (HALILOGLU; BAHAR; ERMAN, 1997) propuseram o modelo de rede gaussiana (GNM). No GNM, a proteína é

representada por uma rede elástica tridimensional, onde os vértices da rede são identificados pelas coordenadas dos  $C_\alpha$  e as molas representam as interações ligantes e não ligantes entre resíduos que estão distantes entre si até um dado limiar. Assume-se ainda que as interações são gaussianas e isotrópicas. No GNM, não há informações sobre as direções dos movimentos. O modelo de rede anisotrópica (ANM) é uma extensão do GNM em que as flutuações são consideradas como sendo anisotrópicas.

Os modelos de análise de modos normais GNM e ANM demonstram ser robustos. Estes modelos têm sido aplicados no estudo do mecanismo e da dinâmica conformacional de sistemas proteicos como: a hemoglobina (MUWAD; PERAHIA, 1996), HIV *transcriptase reverse* (BAHAR e colaboradores., 1999; TEMIS; BAHAR, 2002), *aspartate transcarbamylase* (THOMAS e colaboradores., 1999), GroEL-GroES (KESKIN e colaboradores., 2002), um segmento de *actin* (MING e colaboradores., 2003), *ribosome* (TAMA e colaboradores., 2003; WANG e colaboradores., 2004), RNA polimerase (WYNSBERGHE; LI; CUI, 2004), F1-ATpase (CUI e colaboradores., 2004) e *viral capsids* (RADER; VLAD; BAHAR, 2005).

Recentemente, esses modelos vêm sendo revisados e modificados. Yang e colaboradores (YANG; SONG; JERNIGAM, 2009) propuseram os modelos pfGNM e pfANM, nos quais não há distância de corte entre as interações, ou seja, considera-se que cada resíduo interage com todos os outros resíduos. Além disto, a constante de força da mola virtual que conecta dois resíduos quaisquer é ponderada pelo inverso do quadrado da distância de separação entre eles.

Os objetivos estabelecidos para o desenvolvimento deste trabalho são:

- (i) estudar e compreender os modelos de rede elástica (ENM) propostos na literatura;
- (ii) implementar programas computacionais relativos a estes modelos e validá-los para um número suficientemente grande de proteínas;
- (iii) comprovar ou revelar quais modelos descrevem melhor os fatores de temperatura experimentais;
- (iv) propor novos modelos de rede elástica.

Descrevemos a seguir o que expomos nos próximos capítulos desta dissertação. Apresentamos no capítulo 2 uma revisão bibliográfica sobre a teoria dos modelos de rede elástica. Neste capítulo, o nosso objetivo é mostrar a evolução dos estudos propostos na literatura sobre a análise de modos normais em proteínas. Mais especificamente, enfocamos como as abordagens para a Hamiltoniana evoluíram ao longo dos anos. Para

isto, descrevemos, na seção 2.1, o campo de forças utilizado em simulações de dinâmica molecular de proteínas na década 80 e a sua versão simplificada proposta por Gō e colaboradores. Em seguida, na seção 2.2, demonstramos a abordagem de rede proposta por Tirion e o seu potencial simplificado. Na seção subsequente, apresentamos o modelo GNM e a demonstração da equação que descreve os fatores-B teóricos neste modelo. Mostramos que esta equação é função dos modos normais e das frequências normais de vibração da matriz de Kirchhoff.

Descrevemos nas demais seções deste capítulo o modelo de rede anisotrópica, a abordagem número de contatos ponderados e a versão livre de parâmetro dos modelos GNM e ANM. Apresentamos também as demonstrações das expressões analíticas que descrevem os fatores-B teóricos nestes modelos.

No capítulo 3, expomos, na primeira seção, os procedimentos utilizados para elaborar o conjunto de dados de estruturas protéicas utilizado neste trabalho. Na seção subsequente, descrevemos os algoritmos dos programas computacionais de análise de modos normais com os quais realizamos o estudo comparativo entre os ENM. Este estudo comparativo é apresentado no capítulo 4. Neste estudo, comprovamos que os modelos pfGNM e pfANM descrevem melhor os fatores-B experimentais que os modelos GNM e ANM, conforme mostrado por Yang e colaboradores (YANG; SONG; JERNIGAM, 2009).

Propomos, no capítulo 5, um novo modelo que descreve os fatores-B pseudo-anisotrópicos e mostramos que nosso modelo apresenta uma melhor correlação com os dados experimentais que a versão tradicional ANM.

Por fim, deixamos a apresentação de nossas conclusões e perspectivas para o capítulo 6. Nos apêndices A e B desta dissertação, descrevemos a teoria dos modos normais de vibração. No apêndice B encontram-se as demonstrações das equações introduzidas no apêndice A. Apresentamos no apêndice C uma breve introdução sobre a teoria de grafos. Neste apêndice, definimos a matriz de Kirchhoff e calculamos os elementos da matriz Hessiana no modelo ANM. No apêndice D, inserimos os programas computacionais relativos aos modelos GNM e pfGNM implementados em fortran 77. No apêndice E, apresentamos a lista das proteínas utilizadas nesta dissertação para que os resultados aqui apresentados possam ser reproduzidos.

## 2 Modelos de rede elástica (ENM)

Descreveremos neste capítulo a teoria dos modelos de rede elástica. Inicialmente, apresentaremos a abordagem proposta por Gō e colaboradores (GŌ; NOGUTI; NISHIKAWA, 1993) e, posteriormente, a proposta de Tirion (TIRION, 1996). Na sequência descreveremos o modelo de rede gaussiana e a sua versão anisotrópica. Além destas abordagens, apresentaremos a abordagem estática denominada número de contatos ponderados na qual não é necessário a diagonalização da matriz de Kirchhoff. Por fim, descreveremos o modelo de rede elástica livre de parâmetros, proposto recentemente por Yang e colaboradores (YANG; SONG; JERNIGAM, 2009). Nos ENM, a rede é constituída de resíduos conectados por molas virtuais. As molas representam as interações ligantes e não ligantes entre resíduos conforme ilustrado na Figura 2.

### 2.1 Modelo proposto por Gō

Os primeiros estudos da dinâmica de macromoléculas por meio da análise de modos normais iniciaram-se na década de 80. Em geral, utilizavam-se simulações de dinâmica molecular (MD) envolvendo todos os átomos. Nestas simulações, adotavam-se campos de forças relativamente complexos, por exemplo,

$$\begin{aligned}
 E_p &= \frac{1}{2} \sum_{\text{ligações}} k_b (b - b_0)^2 + \frac{1}{2} \sum_{\text{ângulos}} k_\theta (\theta - \theta_0)^2 \\
 &+ \frac{1}{2} \sum_{\text{torsões}} k_\phi [(1 + \cos(n\phi - \delta))] \\
 &+ \sum_{\text{não ligantes}} \left[ \frac{A}{r^{12}} - \frac{B}{r^6} + \frac{q_1 q_2}{Dr} \right].
 \end{aligned} \tag{2.1}$$

Nesta expressão, o primeiro termo leva em conta a contribuição da energia associada às variações dos comprimentos das ligações, o segundo termo está associado às variações dos ângulos entre as ligações, o terceiro termo inclui as variações dos ângulos torcionais e o último termo leva em conta as interações não ligantes.

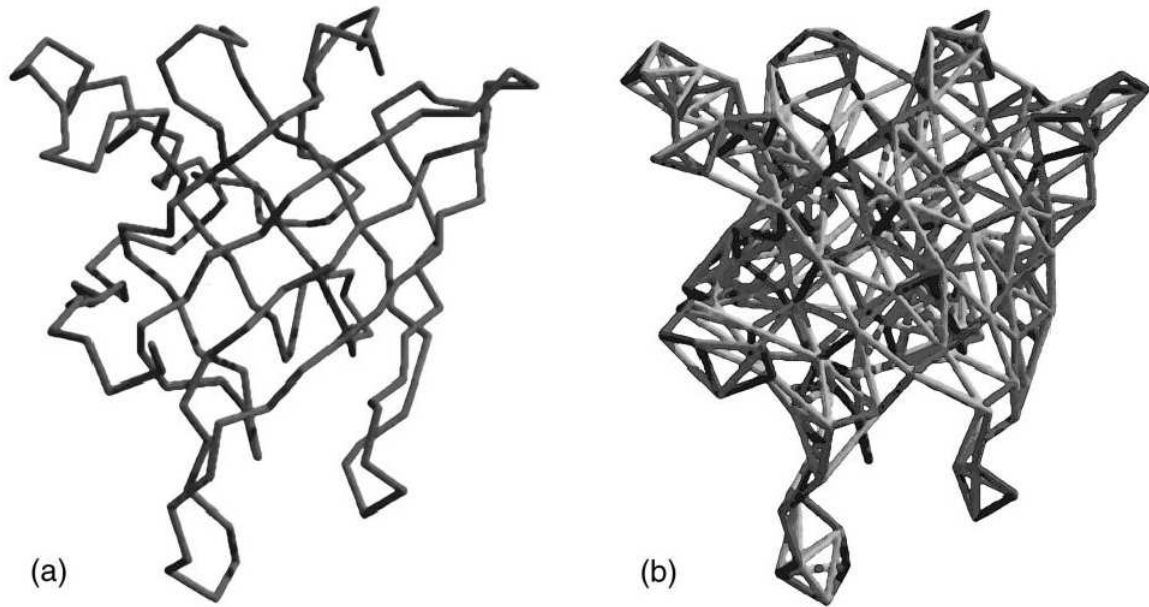


Figura 2: (a) Representação da cadeia linear de resíduos (*backbone*). (b) representação do *backbone* ilustrada pela linha mais escura e da rede elástica tridimensional de resíduos pela linha mais clara. Figura adaptada do artigo (ATILGAN e colaboradores., 2001).

Em MD, para calcular as constantes de força das ligações de uma proteína constituída por  $N$  átomos é necessário fazer  $N$  minimizações da energia. Assim, o custo computacional é diretamente proporcional ao número de átomos e da complexidade do campo de forças escolhido.

Com o intuito de reduzir o custo computacional nas simulações de dinâmica molecular, Gō e colaboradores (Gō; NOGUTI; NISHIKAWA, 1993) propuseram um campo de forças relativamente mais simples do que o apresentado na equação (2.1). Para isso, eles conjecturaram que a dinâmica dos átomos nas proteínas é governada pelas interações de curto alcance,

$$V \approx \sum_i \left[ \frac{k_1}{2} (r_{i,i+1} - r_{i,i+1}^0)^2 + \frac{k_2}{4} (r_{i,i+1} - r_{i,i+1}^0)^4 \right] + \sum_{i,j} 4\epsilon \left( \frac{1}{r_{ij}^6} - \frac{1}{r_{ij}^{12}} \right), \quad (2.2)$$

no qual o primeiro termo abrange as oscilações harmônicas e anarmônicas e o último as interações não ligantes descritas pelo potencial clássico de Lenard-Jonnes. Aqui  $r_{ij}$  é a distância entre os átomos  $i$  e  $j$  e  $r_{ij}^0$  é a correspondente distância na conformação de equilíbrio. Este campo de forças é usualmente conhecido como potencial de Gō, o qual foi amplamente empregado no estudo dos processos de desenovelamento de proteínas nas décadas passadas.

## 2.2 Modelo proposto por Tirion

Inspiraado no modelo proposto por Gō, Tirion (TIRION, 1996) propõe um modelo no qual a minimização da energia é eliminada. Em termos de eficiência computacional, o grande sucesso desse modelo que indubitavelmente acarretou em um avanço nas abordagens *coarse-grained*, que serão descritas posteriormente, foi a escolha da conformação da proteína no estado enovelado obtido do *Protein Data Bank* (PDB) como sendo a conformação de equilíbrio já minimizada. As interações entre pares de resíduos são descritas pelo potencial simplificado com um único parâmetro,

$$E_p(\mathbf{r}_a, \mathbf{r}_b) = \frac{C}{2} (|\mathbf{r}_{a,b}| - |\mathbf{r}_{a,b}^0|)^2, \quad (2.3)$$

em que  $\mathbf{r}_{a,b} \equiv \mathbf{r}_a - \mathbf{r}_b$  é o deslocamento do resíduo  $b$  em relação ao resíduo  $a$ , correspondendo ao alongamento da mola virtual conectando o resíduo  $a$  ao  $b$ . Neste modelo, está implícita a hipótese de que as interações entre os resíduos seguem uma distribuição de probabilidade gaussiana de tal forma que as interações aos pares são representadas por

$$E = \sum_{(a,b)} E_p(\mathbf{r}_a, \mathbf{r}_b). \quad (2.4)$$

Nesta equação, o somatório é restrito aos pares de resíduos cuja distância entre eles é menor que  $R_{\text{vdW}}(a) + R_{\text{vdW}}(b) + R_c$ , onde  $R_{\text{vdW}}(a)$  é o raio de van der Waals do resíduo  $a$ ,  $R_{\text{vdW}}(b)$  o raio de van der Waals do resíduo  $b$  e  $R_c$  o raio de corte delimitando o alcance das interações. Na equação (2.3),  $C$  representa uma constante de força fenomenológica, a qual assume o mesmo valor para todos os pares de resíduos conectados.

## 2.3 Modelo de rede gaussiana (GNM)

Inspirados no modelo de rede elástica proposto por Tirion (TIRION, 1996) e na teoria de rede elástica de polímeros desenvolvida por Flory (FLORY, 1976), Haliloglu e colaboradores (HALILOGLU; BAHAR; ERMAN, 1997) propuseram o modelo de rede gaussiana. A finalidade desse modelo é explorar a dinâmica da rede elástica, definida pela sua estrutura 3D, para obter informações relevantes sobre as flutuações térmicas dos átomos nas proteínas.

No GNM, a proteína é descrita como uma rede elástica tridimensional (ver figura 3a), onde os vértices da rede são identificados pelas coordenadas dos resíduos e as molas representam as interações ligantes e não ligantes entre eles. Duas suposições são feitas em

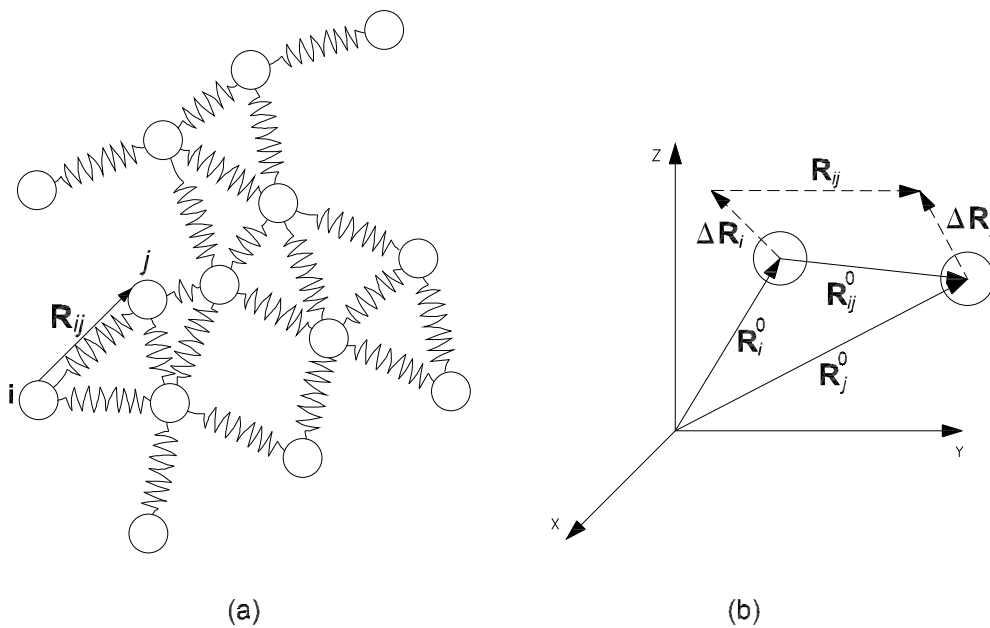


Figura 3: (a) Representação da proteína por uma rede elástica tridimensional, onde os vértices da rede são definidos pelas coordenadas dos resíduos e as molas representam as interações não ligantes entre eles. Neste modelo, considera-se que todas as molas possuem constante de força  $\gamma$ . (b) representação do sistema de referência fixo no laboratório para descrever as posições dos resíduos  $i$  e  $j$ .

relação às flutuações:

- (i) são isotrópicas,
- (ii) seguem uma distribuição de probabilidade gaussiana.

É conveniente estudarmos a rede elástica em relação a um sistema de referência fixo no laboratório (ver figura 3b).

Uma vez escolhido o sistema de referência, a posição de equilíbrio do vértice  $i$  da rede é determinada pelo vetor posição  $\mathbf{R}_i^0$  e a posição instantânea por  $\mathbf{R}(i)$ . As flutuações ou deformações das posições médias podem ser definidas pelo vetor  $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$ . Semelhantemente, as flutuações no vetor distância  $\mathbf{R}_{ij}$ , entre os resíduos  $i$  e  $j$ , são dadas por  $\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i$ .

Tendo em vista as suposições (i) e (ii), podemos escrever a Hamiltoniana da rede de

$N$  resíduos,  $\mathcal{H}_{\text{GNM}}$ , em função das deformações no vetor distância  $\Delta\mathbf{R}_{ij}$  da seguinte forma

$$\begin{aligned}\mathcal{H}_{\text{GNM}} &= \frac{\gamma}{2} \sum_{i=1}^N \sum_{j>i}^N (\Delta\mathbf{R}_{ij})^2 H(R_c - |\mathbf{R}_{ij}^0|) \\ &= \frac{\gamma}{2} \sum_{i=1}^N \sum_{j>i}^N (\mathbf{R}_{ij} - \mathbf{R}_{ij}^0) \cdot (\mathbf{R}_{ij} - \mathbf{R}_{ij}^0) H(R_c - |\mathbf{R}_{ij}^0|)\end{aligned}\quad (2.5)$$

$$= \frac{\gamma}{2} \sum_{i=1}^N \sum_{j>i}^N (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i) \cdot (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i) H(R_c - |\mathbf{R}_{ij}^0|)\quad (2.6)$$

onde  $H(R_c - |\mathbf{R}_{ij}^0|)$  é a função de passo Heaviside (assumindo valor unitário se o argumento for maior do que zero e nulo, caso contrário). A equação acima pode ser expressa na forma quadrática da seguinte forma

$$\mathcal{H}_{\text{GNM}} = \frac{\gamma}{2} \Delta\mathbf{R}^T \mathbf{\Gamma} \Delta\mathbf{R},\quad (2.7)$$

onde  $\Delta\mathbf{R} = [\Delta\mathbf{R}_1, \Delta\mathbf{R}_2, \dots, \Delta\mathbf{R}_N]$  é um hipervetor de dimensão  $N$  cujas componentes são vetores e  $\mathbf{\Gamma}$  é a matriz de kirchhoff (ou matriz conectividade) das interações de contatos de resíduos. Esta matriz tem elementos  $\Gamma_{ij}$ :

$$\Gamma_{ij} = \begin{cases} -1, & \text{se } i \neq j \text{ e } R_{ij} \leq R_c \\ 0, & \text{se } i \neq j \text{ e } R_{ij} > R_c \\ -\sum_{i=1, k \neq i}^N \Gamma_{ik}, & \text{se } i = j. \end{cases}$$

Demonstramos a equação (2.7) ilustrativamente na figura 4.

No GNM, a distribuição de probabilidade de todas as flutuações,  $P(\Delta\mathbf{R})$  é *isotrópica* e *gaussiana*,

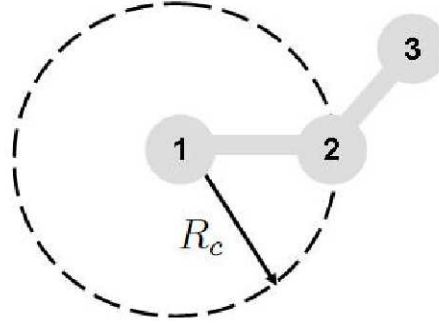
$$P(\Delta\mathbf{R}) = P(\Delta\mathbf{X}, \Delta\mathbf{Y}, \Delta\mathbf{Z}) = p(\Delta\mathbf{X})p(\Delta\mathbf{Y})p(\Delta\mathbf{Z})$$

onde

$$\begin{aligned}p(\Delta\mathbf{X}) &\propto \exp \left\{ -\frac{\gamma}{2k_B T} \Delta\mathbf{X}^T \mathbf{\Gamma} \Delta\mathbf{X} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \Delta\mathbf{X}^T \left( \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right)^{-1} \Delta\mathbf{X} \right] \right\},\end{aligned}$$

$k_B$  é a constante de Boltzmann e  $T$  representa a temperatura absoluta. Analogamente, temos expressões semelhantes para  $p(\Delta\mathbf{Y})$  e  $p(\Delta\mathbf{Z})$ . A notação  $\Delta\mathbf{X}$  corresponde a  $[\Delta X_1, \Delta X_2, \dots, \Delta X_i, \dots, \Delta X_N]$  e é também uma variável aleatória gaussiana multidimen-





$$\begin{aligned}
\mathcal{H}_{\text{GNM}} &= \frac{\gamma}{2} [(\Delta\mathbf{R}_{12})^2 + (\Delta\mathbf{R}_{23})^2] \\
&= \frac{\gamma}{2} [(\Delta\mathbf{R}_2 - \Delta\mathbf{R}_1)^2 + (\Delta\mathbf{R}_3 - \Delta\mathbf{R}_2)^2] \\
&= \frac{\gamma}{2} \underbrace{\begin{bmatrix} \Delta\mathbf{R}_1 & \Delta\mathbf{R}_2 & \Delta\mathbf{R}_3 \end{bmatrix}}_{\text{matriz de kirchhoff } \mathbf{\Gamma}} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{R}_1 \\ \Delta\mathbf{R}_2 \\ \Delta\mathbf{R}_3 \end{bmatrix}
\end{aligned}$$

Figura 4: Energia potencial de um tripeptídeo. Como a distância entre o resíduo 1 e o 3 é maior que  $R_c$ , desconsidera-se esta interação. A equação final da Hamiltoniana  $\mathcal{H}_{\text{GNM}}$  na forma matricial é semelhante a equação (2.7).

sional com média zero e covariância  $(\frac{k_B T}{\gamma})\mathbf{\Gamma}^{-1}$  conforme a definição geral a seguir,

$$W(\mathbf{x}, \mu, \Xi) = \frac{1}{(2\pi)^N |\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^T \Xi^{-1} (\mathbf{x} - \mu) \right\},$$

para a função distribuição de probabilidade gaussiana multidimensional associada com um vetor  $\mathbf{x}$   $N$ -dimensional tendo valor médio  $\mu$  e matriz covariância  $\Xi$ . Aqui, o termo no denominador,  $(2\pi)^N |\Xi|^{1/2}$ , é a função de partição que garante a normalização de  $W(\mathbf{x}, \mu, \Xi)$  e  $|\Xi|$  é o determinante de  $\Xi$ . Semelhantemente, a distribuição de probabilidade normalizada  $p(\Delta\mathbf{X})$  é

$$p(\Delta\mathbf{X}) = \frac{1}{Z_{\mathbf{X}}} \exp \left\{ -\frac{1}{2} \left[ \Delta\mathbf{X}^T \left( \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right)^{-1} \Delta\mathbf{X} \right] \right\},$$

onde  $Z_{\mathbf{X}}$  é a função de partição dada por

$$Z_{\mathbf{X}} = \int \exp \left\{ -\frac{1}{2} \left[ \Delta\mathbf{X}^T \left( \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right)^{-1} \Delta\mathbf{X} \right] \right\} d\Delta\mathbf{X} = (2\pi)^{N/2} \left| \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right|^{1/2}.$$

Sendo as flutuações isotrópicas, podemos escrever a função de partição configuracional

total como

$$Z = Z_{\mathbf{x}}Z_{\mathbf{y}}Z_{\mathbf{z}} = (2\pi)^{3N/2} \left| \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right|^{3/2}.$$

Uma vez determinada a função de partição, temos então toda a informação necessária para calcularmos as grandezas médias deste sistema. Assim, a autocorrelação associada ao resíduo  $i$  pode ser dada por

$$\langle \Delta \mathbf{X}_i^T \Delta \mathbf{X}_i \rangle = \frac{\int \Delta \mathbf{X}_i^T \Delta \mathbf{X}_i \exp \left\{ -\frac{1}{2} \left[ \Delta \mathbf{X}^T \left( \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right)^{-1} \Delta \mathbf{X} \right] \right\} d\Delta \mathbf{X}}{\int \exp \left\{ -\frac{1}{2} \left[ \Delta \mathbf{X}^T \left( \frac{k_B T}{\gamma} \mathbf{\Gamma}^{-1} \right)^{-1} \Delta \mathbf{X} \right] \right\} d\Delta \mathbf{X}}.$$

Utilizando a identidade

$$\frac{\partial}{\partial \gamma} \ln Z = \frac{1}{Z} \frac{\partial Z}{\partial \gamma},$$

obtemos

$$\langle \Delta \mathbf{X}_i^T \Delta \mathbf{X}_i \rangle = -\frac{\partial}{\partial \gamma} \ln Z.$$

Como as flutuações são isotrópicas, podemos escrever

$$\langle \Delta \mathbf{X}_i^T \Delta \mathbf{X}_i \rangle = \langle \Delta \mathbf{Y}_i^T \Delta \mathbf{Y}_i \rangle = \langle \Delta \mathbf{Z}_i^T \Delta \mathbf{Z}_i \rangle = \frac{1}{3} \langle \Delta \mathbf{R}_i^T \Delta \mathbf{R}_i \rangle.$$

Portanto, os deslocamentos quadráticos médios podem ser escritos como

$$\langle \Delta \mathbf{R}_i^T \Delta \mathbf{R}_i \rangle = \frac{3k_B T}{\gamma} (\mathbf{\Gamma}^{-1})_{ii} \quad (2.8)$$

e

$$\langle \Delta \mathbf{R}_i^T \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} (\mathbf{\Gamma}^{-1})_{ij}. \quad (2.9)$$

O determinante da matriz de Kirchhoff  $\mathbf{\Gamma}$  é nulo, portanto, a princípio ela não é inversível (matriz singular). Entretanto, ela pode ser decomposta em termos de uma matriz  $\mathbf{U}$ , cujas colunas são os autovetores  $\mathbf{u}_i$ , e da matriz diagonal  $\mathbf{\Lambda}$ , com autovalores  $\lambda_i$ , da seguinte forma

$$\mathbf{\Gamma} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (2.10)$$

onde os autovalores da matriz de Kirchhoff representam as frequências dos modos normais individuais enquanto que os autovetores definem as direções dos modos. Por conveniência os autovalores são ordenados em ordem crescente. O primeiro autovalor é nulo,  $\lambda_1 = 0$ , o qual corresponde à translação da proteína como um todo. Ignorando o primeiro autovalor, a matriz pseudo-inversa de Kirchhoff pode ser escrita como

$$\mathbf{\Gamma}^{-1} = \sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T. \quad (2.11)$$

Consequentemente, as correlações cruzadas e as autocorrelações associadas às flutuações dos resíduos  $i$  e  $j$  podem ser expressas respectivamente como:

$$\langle \Delta \mathbf{R}_i^T \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} \sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_{ik} \mathbf{u}_{jk}^T, \quad (2.12)$$

$$\langle \Delta \mathbf{R}_i^T \Delta \mathbf{R}_i \rangle = \frac{3k_B T}{\gamma} \sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_{ik}^2. \quad (2.13)$$

O fator-B, o qual mede a mobilidade dos resíduos, é definido como

$$B_i = \frac{8\pi^2}{3} \langle (\Delta \mathbf{R}_i)^2 \rangle. \quad (2.14)$$

Utilizando a equação (2.13), podemos escrevê-lo como

$$B_i = \frac{8\pi^2 k_B T}{\gamma} \sum_{k=2}^N \lambda_k^{-1} \mathbf{u}_{ik}^2. \quad (2.15)$$

Assim, o fator-B pode ser calculado teoricamente e comparado com o experimental, conforme ilustrado na figura 5.

Na equação (2.15),  $\gamma$  é uma constante a ser determinada. Usualmente, ela é obtida por meio da normalização dos fatores-B teóricos com os experimentais. Razão pela qual nos referimos a ela como constante de normalização. Fazendo  $B_i = B_i^{\text{GNM}}$ , podemos obter o seu valor (em unidades de  $K_B T$ ) através da expressão

$$\gamma = \frac{\sum_{i=1}^N B_i^{\text{GNM}}}{\sum_{i=1}^N B_i^{\text{experimental}}}.$$

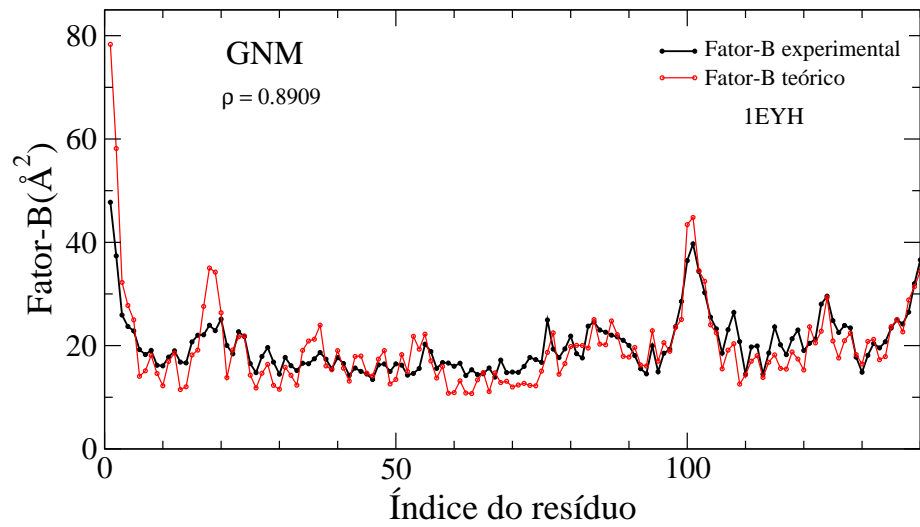


Figura 5: Comparação entre os espectros do fator de temperatura teórico determinado pelo GNM e o experimental.

Os pontos nas curvas ilustradas na figura 5 para proteína 1EYH estão associados ao fator-B do  $C_\alpha$  no aminoácido correspondente. Observa-se que as curvas teórica e experimental estão bem correlacionadas, apresentando comportamentos similares. O coeficiente de correlação de Pearson entre elas é  $\rho = 0.8909$ .

## 2.4 Modelo de rede gaussiana (ANM)

No GNM é possível calcular os valores das magnitudes das flutuações individuais dos resíduos, porém não é possível extrair informações sobre as direções dessas flutuações. Entretanto, na realidade as flutuações são em geral anisotrópicas. Acessar as direções preferenciais dos movimentos coletivos dos resíduos é importante porque estas podem ser diretamente relevantes no desempenho de funções biológicas da proteína. Apresentaremos aqui uma extensão do GNM denominada modelo de rede anisotrópica (ANM) introduzida por Altigan e colaboradores (ATILGAN e colaboradores., 2001). No ANM assume-se que as flutuações dos resíduos são anisotrópicas. A matriz Hessiana  $\mathbf{H}$  para uma rede de  $N$  vértices é uma matriz  $3N \times 3N$  composta de  $N \times N$  super elementos,  $\mathbf{H}_{ij}(i \neq j)$ , definidos da seguinte forma:

$$\mathbf{H}_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix}, \quad (2.16)$$

na qual a energia potencial é definida por

$$V = \frac{\gamma}{2} \sum_i \sum_{j>i} (R_{ij} - R_{ij}^0)^2 H(R_c - R_{ij}^0), \quad (2.17)$$

onde  $\gamma$  é uma constante de normalização,  $R_{ij}$  é a distância entre as posições instantâneas dos resíduos  $i$  e  $j$  e  $R_{ij}^0$  a distância entre eles na configuração de equilíbrio.

Das equações (2.16) e (2.17) podemos escrever os super elementos  $\mathbf{H}_{ij}(i \neq j)$  da seguinte forma (ver apêndice C.3),

$$\mathbf{H}_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix} H(R_c - R_{ij}^0), \quad (2.18)$$

em que  $X_{ij}$ ,  $Y_{ij}$  e  $Z_{ij}$  são as coordenadas associadas à distância  $R_{ij}$ . Por sua vez, os  $N \times N$  super elementos da diagonal são definidos como

$$\mathbf{H}_{ii} = -\sum_{j \neq i} \mathbf{H}_{ij}.$$

Analogamente ao desenvolvimento feito no GNM, a matriz pseudo-inversa, ou inversa generalizada, pode ser decomposta em função dos autovetores e autovalores da seguinte

forma

$$\mathbf{H}^{-1} = \sum_{k=7}^{3N} \frac{1}{\lambda_k} \mathbf{u}_k \mathbf{u}_k^T, \quad (2.19)$$

em que os 6 primeiros autovalores são nulos e estão associados à translação e rotação da proteína (corpo rígido) como um todo. Analogamente ao fator de temperatura do GNM, o fator de temperatura anisotrópico teórico é dado por

$$B_i^{ANM} = \frac{8\pi^2 k_B T}{\gamma} Tr(\mathbf{H}^{-1})_{ii}.$$

O fator de temperatura anisotrópico experimental é representado por uma matriz simétrica  $B^{\text{aniso}}$  de dimensão  $3 \times 3$ . Tal quantidade contém informações sobre as magnitudes e as direções preferenciais das flutuações térmicas e é escrita como:

$$B^{\text{aniso}} = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}.$$

Fisicamente, o fator de temperatura anisotrópico descreve a função densidade de probabilidade associada às flutuações térmicas residuais por meio de uma função gaussiana tridimensional. Fixando-se um valor de probabilidade, a distribuição é elipsoidal e tem uma preferência direcional.

### 2.4.1 Comparando fatores-B anisotrópicos com dados experimentais

Na seção (2.3), a comparação entre os fatores-B isotrópicos teóricos e os experimentais foi feita mediante a correlação entre eles. Ou seja, pela correlação entre dois vetores. Entretanto, os fatores de temperatura anisotrópicos são matrizes, portanto, não podemos utilizar esta abordagem.

Cada matriz representa uma função distribuição de probabilidade tridimensional, a qual podemos visualizar como sendo um elipsóide. Dessa forma, a comparação entre os fatores de temperatura anisotrópicos pode ser feita via comparação entre dois elipsóides em termos das suas formas e orientações. Para isto, devemos primeiramente diagonalizar as matrizes. Os seus autovetores nos fornecerão a magnitude e as formas das flutuações. As direções preferenciais das flutuações serão ditadas pelos autovetores (YANG; SONG; JERNIGAN, 2009). Os três autovetores de uma matriz formam um sistema ortogonal. Podemos comparar a orientação dos dois elipsóides pelo ângulo formado entre os dois sistemas ortogonais.

## 2.5 Número de contatos ponderados (WCN)

Nesta abordagem, proposta por Lin e colaboradores (LIN e colaboradores., 2008), o número de contatos (NC) de um vértice qualquer da rede é definido pelo número de vizinhos com os quais ele está conectado. Novamente, consideramos que dois vizinhos estão conectados se a distância entre eles estiver até um dado limiar  $R_c$ . Segundo a teoria dos grafos, podemos representar o NC do vértice  $i$  por meio do seu grau  $\nu_i$ , o qual é descrito pelo elemento  $\Gamma_{ii}$  da matriz de Kirchhoff,

$$\begin{aligned}\nu_i &= \Gamma_{ii} \\ &= \sum_{j \neq i}^N H(R_c - R_{ij}).\end{aligned}\quad (2.20)$$

No WCN, define-se o NC dependente da distância da seguinte forma

$$\nu'_i = \sum_{j \neq i}^N H(R_c - R_{ij}) / R_{ij}^2. \quad (2.21)$$

Nesta equação, a soma é ponderada pelo recíproco do quadrado da distância entre pares de  $C_\alpha$ . Como o termo  $1/R_{ij}^2$  decai rapidamente para grandes separações, então podemos simplificá-la para

$$\nu'_i = \sum_{j \neq i}^N \frac{1}{R_{ij}^2}. \quad (2.22)$$

O perfil WCN do fator-B de uma proteína com  $N$  resíduos é definido por

$$\mathbf{w} = (\omega_1, \omega_2, \dots, \omega_N), \quad (2.23)$$

onde  $\omega_i = 1/\nu'_i$ . Podemos interpretar fisicamente a equação (2.23) recorrendo ao problema clássico de dois osciladores acoplados representados pelos carbonos  $\alpha$   $i$  e  $j$  conectados por uma mola virtual. As flutuações  $R_{ij}$  destes osciladores são inversamente proporcionais à constante de força associada à mola. Assim, podemos obter o número de contatos ponderados  $\nu'_i$  por meio do elemento  $\Gamma_{ii}$  da matriz de Kirchhoff,

$$\Gamma_{ij}^{\text{WCN}} = \begin{cases} R_{ij}^{-2}, & \text{se } i \neq j \\ -\sum_{i,j \neq i} \Gamma_{ij}, & \text{se } i = j. \end{cases} \quad (2.24)$$

Consequentemente, o perfil do fator-B teórico associado ao  $i$ -ésimo resíduo é dado por  $\omega_i = 1/\Gamma_{ii}$ . Desta forma, o coeficiente de correlação entre o fator-B experimental e o

teórico relativo ao WCN pode ser obtido por meio da correlação entre dois vetores  $\mathbf{b}$  e  $\mathbf{w}$ , sendo  $\mathbf{b}$  o perfil do fator-B experimental descrito por

$$\mathbf{b} = (b_1, b_2 \dots b_N), \quad (2.25)$$

onde  $b_i$  é o fator-B experimental do  $i$ -ésimo  $C_\alpha$  da estrutura nativa obtida no PDB.

Assim, nesta abordagem não é necessário diagonalizar a matriz de Kirchhoff. As flutuações térmicas são calculadas diretamente da equação (2.23). A eliminação dessa etapa resulta em uma diminuição significativa das operações de ponto flutuante, reduzindo o custo computacional. Lin e colaboradores (LIN e colaboradores., 2008) mostraram que os resultados obtidos pelo WCN são competitivos com os determinados pelo GNM.



## 2.6 Modelos de rede elástica livres de parâmetro (pfGNM e pfANM)

Recentemente, Yang e colaboradores (YANG; SONG; JERNIGAM, 2009) introduziram as seguintes modificações nos modelos GNM e ANM:

i) não há distância de corte delimitando as interações entre os resíduos. Assim, considera-se que cada  $C_\alpha$  interage com todos os outros  $C_\alpha$  da rede;

ii) os elementos  $i$  e  $j$  da matriz de Kirchhoff no GNM e da matriz Hessiana no ANM são ponderados pelo inverso do quadrado da distância  $R_{ij}$  observada entre eles na conformação nativa. Consequentemente, podemos escrever

$$\mathbf{\Gamma}_{ij}^{pf} = \begin{cases} R_{ij}^{-2} & \text{se } i \neq j \\ -\sum_{i,j \neq i} \mathbf{\Gamma}_{ij}, & \text{se } i = j \end{cases}$$

e

$$H_{ij}^{pf} = H_{ij} R_{ij}^{-2}.$$

Substituindo estas matrizes nos modelos GNM e ANM descritos nas seções anteriores e mantendo o mesmo procedimento de obtenção do fator-B teórico, Yang e colaboradores (YANG; SONG; JERNIGAM, 2009) propuseram as versões pfGNM e pfANM. Utilizando estas versões, eles mostraram que os valores do coeficiente de correlação entre o fator-B experimental e o teórico apresentam valores significativamente melhores que nas suas versões originais. Comprovamos estas afirmações nas tabelas 1 e 2.

Fisicamente, estes modelos são mais plausíveis que os originais pois são consideradas as interações entre todos os resíduos. Além disto, elas são ponderadas pelo inverso do quadrado da distância, atribuindo pesos menores para as interações a longa distância e maiores para as interações entre primeiros vizinhos. Podemos fazer uma analogia com a lei de Coulomb e da gravitação, nas quais as interações são ponderadas pelo recíproco do quadrado das distâncias e possuem uma equação analítica simples, guardando as características essenciais do sistema.

## 3 Métodos

Deixamos para este capítulo a apresentação dos procedimentos utilizados na elaboração do nosso banco de dados de proteínas e dos métodos computacionais relativos aos modelos de rede elástica. Como os métodos GNM, ANM, pfGNM, pfANM e WCN já foram descritos do ponto de vista teórico no capítulo anterior, descrevemos brevemente, na seção 3.2, estes métodos do ponto de vista computacional.

Inicialmente, apresentamos o procedimento adotado para elaborar o banco de dados de proteínas. Posteriormente, descrevemos em detalhes o algoritmo computacional relativo ao modelo GNM, o qual efetua numericamente o cálculo do espectro dos fatores-B teóricos e da correlação entre este e o espectro dos fatores-B experimentais. Por fim, apresentamos os nomes das subrotinas e bibliotecas empregadas para efetuar a diagonalização da matriz de Kirchhoff e da matriz Hessiana.

### 3.1 Banco de dados de proteínas

Utilizamos neste trabalho como objeto de estudo 500 estruturas protéicas coletadas no *Protein Data Bank* (PDB); um repositório de dados de estruturas 3D de macromoléculas biológicas como proteínas e nucleotídeos, de domínio público, disponível no endereço <http://www.pdb.org/pdb/home/home.do>. Com o auxílio da ferramenta de busca avançada encontrada nesta página, coletamos as estruturas que satisfazem as condições:

- são classificadas de acordo com SCOP em *all- $\alpha$* , *all- $\beta$* ,  *$\alpha+\beta$*  e  *$\alpha/\beta$* ;
- possuem sequências semelhantes com identidade de no máximo 70%;
- possuem tamanhos entre 100 e 1000 resíduos;
- são resolvidas experimentalmente por cristalografia de raios-X, com resolução entre 0 e 2 angstroms.

A listas das proteínas selecionadas encontram-se no apêndice E.

## 3.2 Programas computacionais de análise de modos normais

De acordo com a equação (2.15), os fatores-B teóricos são funções dos modos normais e das frequências normais de vibração da matriz de Kirchhoff. Portanto, para obtê-los numericamente e, conseqüentemente, a correlação entre estes e os fatores-B experimentais, devemos:

- (i) construir a matriz de Kirchhoff, a qual pode ser facilmente obtida pelo cálculo da distância entre os contatos nativos da estrutura do PDB;
- (ii) diagonalizar esta matriz, isto é, obter os seus autovetores e autovalores;
- (iii) determinar os valores dos fatores-B teóricos utilizando os valores dos autovetores e autovalores calculados no passo anterior;
- (iv) calcular o coeficiente de correlação de Pearson  $c$  entre os fatores-B teóricos e experimentais, representados pelos vetores  $B_{\text{teórico}}$  e  $B_{\text{experimental}}$ , pela equação

$$c = \frac{\sum_{i=1}^N (B_{\text{teórico}} - \bar{B}_{\text{teórico}}) (B_{\text{experimental}} - \bar{B}_{\text{experimental}})}{\sqrt{\sum_{i=1}^N (B_{\text{teórico}} - \bar{B}_{\text{teórico}})^2} \sqrt{\sum_{i=1}^N (B_{\text{experimental}} - \bar{B}_{\text{experimental}})^2}}.$$

O valor  $c = 1$  representa uma perfeita correlação entre estes dois vetores, enquanto que  $c = -1$  representa uma perfeita anticorrelação. Quanto a sua intensidade, os valores de  $c$  podem ser classificados como:

- de 0 a 0,3 temos uma fraca correlação;
- de 0,3 a 0,7 uma correlação moderada;
- a partir de 0,7 uma forte correlação.

Numericamente, implementamos o algoritmo acima em *fortran 77* e utilizamos o sistema operacional Linux.

Em (i), para construir a matriz de Kirchhoff, é necessário ler as coordenadas  $x_i, y_i$  e  $z_i$  dos  $N$  resíduos e calcular a distância entre todos os pares de resíduos nativos que estão dentro de uma casca esférica de raio  $R_c$ . Para isso, desenvolvemos uma subrotina em *fortran* da qual extraímos informações, como as coordenadas e o fator de estrutura de cada resíduo, a partir dos arquivos no formato PDB.

Para diagonalizar a matriz de Kirchhoff, no passo (ii) e, conseqüentemente, obter o espectro dos fatores-B teóricos em (iv), utilizamos as bibliotecas otimizadas *Numerical Recipes* (PRESS e colaboradores., 1992), *lapack* e *blzpack* (MARQUES, 1995), para extrair os autovetores e autovalores de pequenas e grandes matrizes de conectividade.

Em (iii), para avaliar o coeficiente de correlação entre os vetores  $B_{\text{experimental}}$  e  $B_{\text{teórico}}$ , foi utilizado a subrotina *pearsn* da biblioteca do livro *Numerical Recipes*

Os programas computacionais dos modelos ANM, pfANM e pfGNM são análogos ao do GNM. O tipo de matriz que descreve as interações entre os contatos nativos é o que diferencia o algoritmo de um modelo do outro. Por exemplo, no modelo ANM, em (i), ao invés de se implementar a matriz de Kirchhoff  $N \times N$ , implementa-se a matriz Hessiana  $3N \times 3N$ , de acordo com a equação (2.18). Já no modelo WCN, o passo (ii) é eliminado, visto que o fator-B neste modelo independe dos modos e das frequências normais da matriz de Kirchhoff. Em (i), implenta-se a matriz de Kirchhoff segundo a equação (2.24) e no passo (iii), a equação dos fatores-B teóricos, a qual depende apenas do número de contatos nativos. Inserimos no apêndice D os códigos dos programas referentes aos modelos GNM e pfGNM.

## 4 Análise comparativa entre os modelos de rede elástica

A tabela 1 mostra os valores médios dos coeficientes de correlação  $\bar{c}$  entre os fatores-B experimentais e os calculados teoricamente pelos modelos GNM, ANM, pfANM, pfGNM e WCN. O valor de  $\bar{c}$  corresponde a uma média sobre 125 estruturas protéicas de cada classe da classificação SCOP: *all- $\alpha$* , *all- $\beta$* ,  *$\alpha+\beta$*  e  *$\alpha/\beta$* . Nesta tabela, são mostrados apenas os valores médios dos coeficientes de correlação. Isto nos diz o quão bem cada modelo descreve o comportamento do espectro experimental do fator-B, em média. Entretanto, não sabemos como estes valores estão distribuídos em cada classe SCOP. Informações mais detalhadas a respeito da distribuição destes coeficientes de correlação são mostradas nos histogramas das figuras 6, 7, 8, 9 e 10. Nestes histogramas, observa-se que, em geral, todos os modelos apresentam ótimos resultados para um fração relativamente pequena de proteínas. Por exemplo, no modelo pfGNM (figura 9), 21,6% das proteínas da classe *all- $\alpha$*  apresentam valores de  $c$  entre 0,7 e 0,8, e apenas 8% entre 0,8 e 0,9. Para todo o conjunto de estruturas protéicas, nota-se que todos os modelos não descrevem bem os dados experimentais, pois exibem valores médios de  $\bar{c}$  em torno de 0,6, valores que correspondem a uma correlação moderada.

Tabela 1: Coeficiente de correlação médio  $\bar{c}$  calculado pelos modelos GNM, ANM, pfANM, pfGNM e WCN para as estruturas distribuídas de acordo com a classificação SCOP. O valor do raio de corte é representado no subscrito de  $\bar{c}$  para os modelos que utilizam este parâmetro.

SCOP classes	GNM $\bar{c}_{7.3}$	ANM $\bar{c}_{13}$	pfANM $\bar{c}$	pfGNM $\bar{c}$	WCN $\bar{c}$
<i>all-<math>\alpha</math></i>	0,56 $\pm$ 0,18	0,49 $\pm$ 0,19	0,57 $\pm$ 0,18	0,60 $\pm$ 0,16	0,61 $\pm$ 0,15
<i>all-<math>\beta</math></i>	0,58 $\pm$ 0,17	0,53 $\pm$ 0,17	0,59 $\pm$ 0,16	0,62 $\pm$ 0,15	0,62 $\pm$ 0,15
<i><math>\alpha+\beta</math></i>	0,58 $\pm$ 0,15	0,50 $\pm$ 0,17	0,58 $\pm$ 0,14	0,62 $\pm$ 0,12	0,62 $\pm$ 0,12
<i><math>\alpha/\beta</math></i>	0,59 $\pm$ 0,13	0,52 $\pm$ 0,17	0,60 $\pm$ 0,14	0,63 $\pm$ 0,12	0,63 $\pm$ 0,12

Além da análise visual dos histogramas, utilizamos uma métrica a qual é definida pela proporção de estruturas protéicas que obtiveram valores de  $c$  maiores que 0,5. A tabela 2 exibe os valores destas métricas para os modelos analisados.

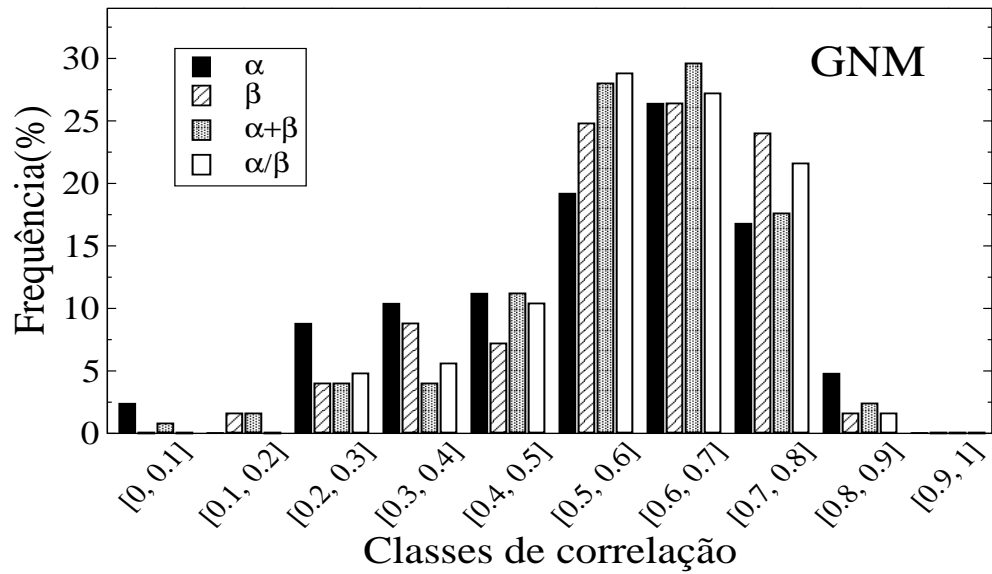


Figura 6: Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo GNM para proteínas das classes *all- $\alpha$* , *all- $\beta$* ,  *$\alpha + \beta$*  e  *$\alpha / \beta$* .

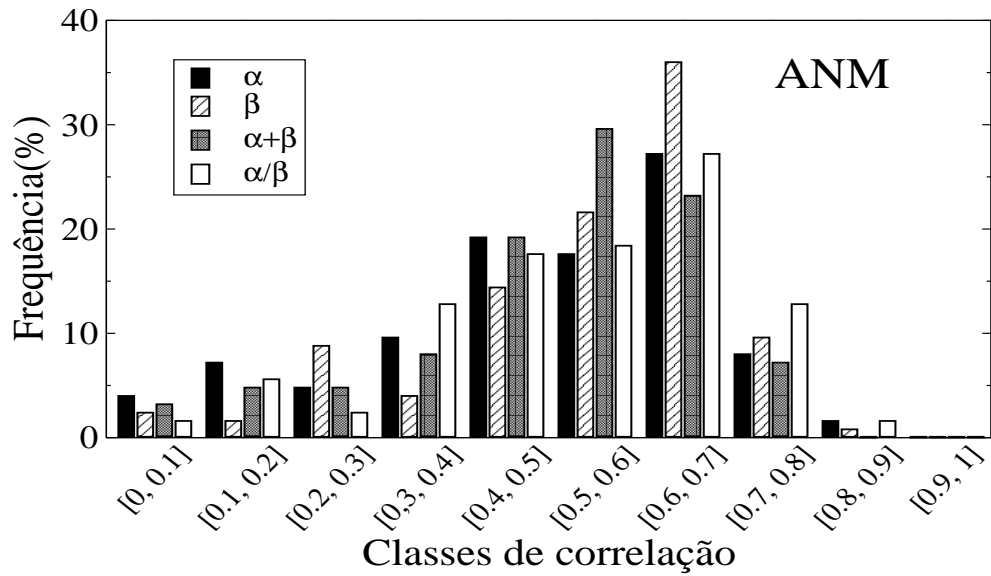


Figura 7: Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo ANM para proteínas das classes *all- $\alpha$* , *all- $\beta$* ,  $\alpha + \beta$  e  $\alpha/\beta$ .

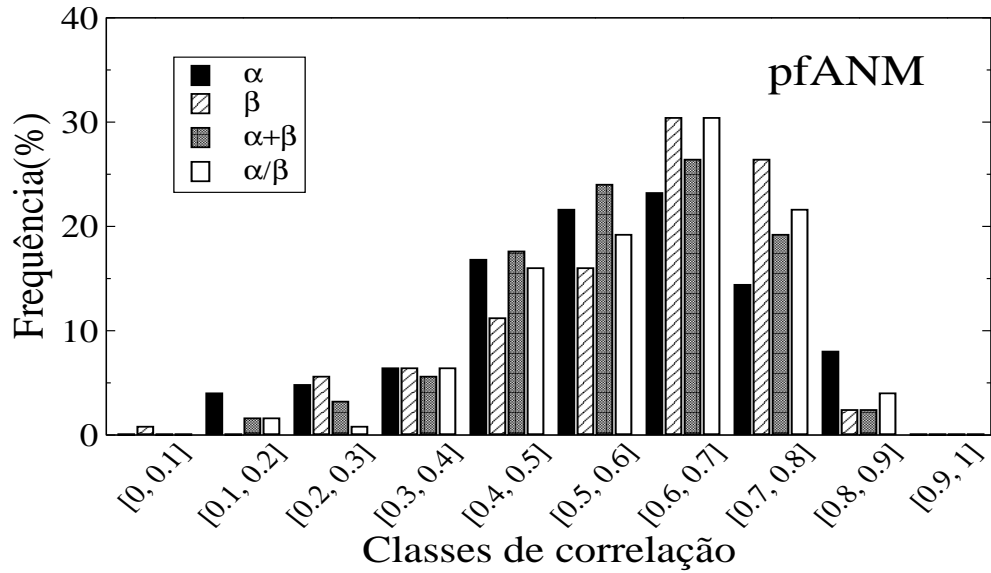


Figura 8: Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo pfANM para proteínas das *all- $\alpha$* , *all- $\beta$* ,  $\alpha + \beta$  e  $\alpha/\beta$ .

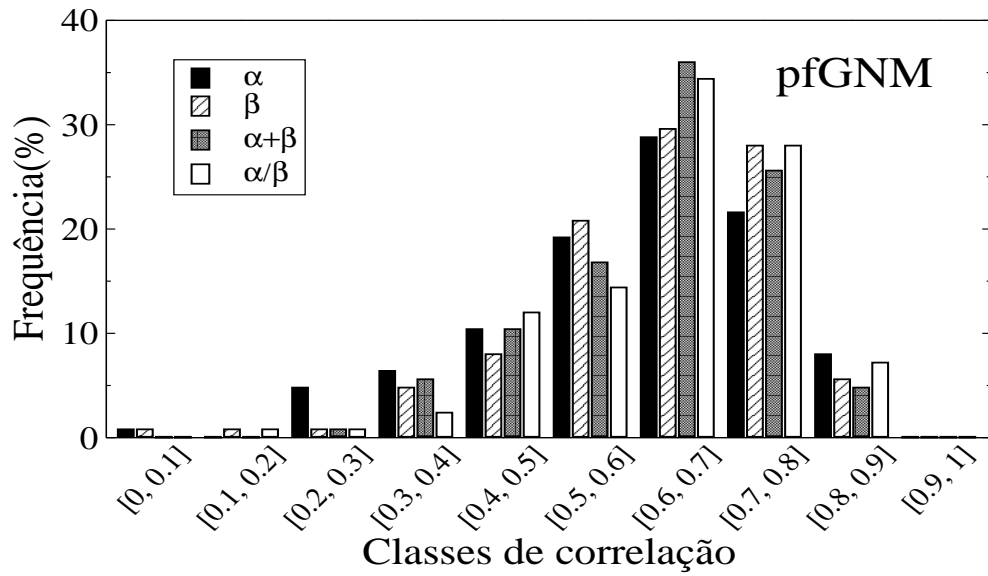


Figura 9: Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo pfGNM para proteínas das classes  $all-\alpha$ ,  $all-\beta$ ,  $\alpha + \beta$  e  $\alpha/\beta$ .

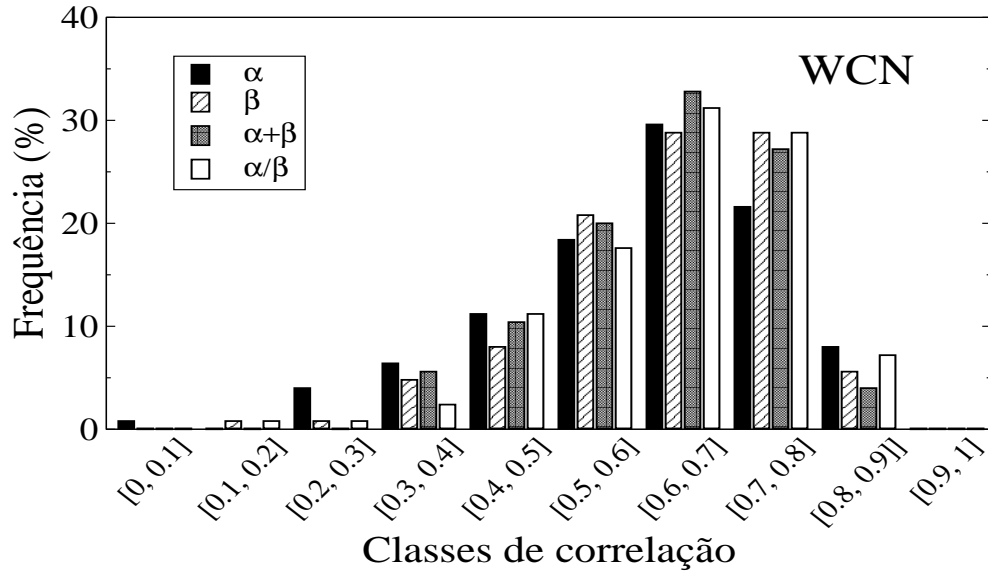


Figura 10: Distribuição do coeficiente de correlação entre o fator-B experimental e o calculado teoricamente pelo WCN para proteínas das classes  $all-\alpha$ ,  $all-\beta$ ,  $\alpha + \beta$  e  $\alpha/\beta$ .



Tabela 2: Proporção das proteínas que obtiveram valores de  $c$  maiores que 0,5.

SCOP classes	GNM	ANM	pfANM	pfGNM	WCN
<i>all-<math>\alpha</math></i>	67,2	54,4	77,6	67,2	77,6
<i>all-<math>\beta</math></i>	76,8	68,0	84,0	75,2	84,0
<i><math>\alpha+\beta</math></i>	77,6	60,0	83,2	70,2	84,0
<i><math>\alpha/\beta</math></i>	79,2	60,0	84,0	75,2	84,8

Na tabela acima, observa-se que o modelo ANM fornece resultados, em média, inferiores aos outros modelos. Fato que também é observado na tabela 1. Os demais modelos fornecem resultados aproximadamente equivalentes, em média. Nota-se uma melhora significativa quando passamos do modelo ANM para o modelo pfANM. Comprovando que os modelos livres de parâmetros descrevem melhor os fatores-B experimentais.

## 5 Novos modelos

### 5.1 Modelo AWCN

A tabela 1 mostra que os valores de  $\bar{c}$  calculados pelos modelos pfGNM e WCN, para as proteínas da classe *all- $\alpha$* , estão muito próximos entre si, enquanto que para as outras classes são idênticos. Este resultado é razoável porque as matrizes de Kirchhoff para ambos os modelos são idênticas, havendo somente diferença nos procedimentos de obtenção do fator-B.

Como no modelo WCN não é necessário diagonalizar matrizes, isto o torna mais eficiente computacionalmente do que os outros modelos, principalmente, mais eficiente que o modelo ANM, no qual é necessário diagonalizar uma matriz de dimensão  $3N \times 3N$ .

Tendo em vista a semelhança entre os resultados obtidos pelos modelos WCN e pfGNM e a eficácia do WCN, por analogia, levantamos a seguinte tese: Dado que os valores de  $\bar{c}$  obtidos pelos modelos pfGNM e WCN foram muito próximos entre si, então uma versão anisotrópica do WCN apresentará resultados muito próximos ou superiores que os determinados pelo ANM. Para testar nossa tese, introduzimos e validamos a abordagem WCN anisotrópica (AWCN), a qual descreveremos a seguir.

Definimos o perfil WCN anisotrópico do fator-B de uma proteína com  $N$  resíduos por

$$\mathbf{w} = (\omega_1, \omega_2, \dots, \omega_N), \quad (5.1)$$

onde  $\omega = 1/\nu'_{\text{ANM}}$ . Analogamente, definimos  $\omega_i$  em função da matriz Hessiana da seguinte forma:

$$\omega_i = 1/H^{pf}(3i-2, 3i-2) + 1/H^{pf}(3i-1, 3i-1) + 1/H^{pf}(3i, 3i). \quad (5.2)$$

Os três termos na equação acima levam em conta as contribuições das flutuações térmicas (deslocamento quadrático médio pseudo-anisotrópico) nas direções  $x$ ,  $y$  e  $z$ . Então, o coeficiente de correlação entre o fator-B experimental e o teórico relativo ao AWCN pode ser obtido por meio da correlação entre os vetores  $\mathbf{b}$  e  $\mathbf{w}$ . Sendo  $\mathbf{b}$  o perfil do fator-B

Tabela 3: Coeficiente de correlação médio para cada modelo apresentado e para as estruturas distribuídas de acordo com a classificação SCOP.

Modelos	$all-\alpha$	$all-\beta$	$\alpha+\beta$	$\alpha/\beta$
GNM	$0,56 \pm 0,18$	$0,58 \pm 0,17$	$0,58 \pm 0,15$	$0,59 \pm 0,13$
ANM	$0,49 \pm 0,19$	$0,53 \pm 0,17$	$0,50 \pm 0,17$	$0,52 \pm 0,17$
pfANM	$0,57 \pm 0,18$	$0,59 \pm 0,16$	$0,58 \pm 0,14$	$0,60 \pm 0,14$
pfGNM	$0,60 \pm 0,16$	$0,62 \pm 0,15$	$0,62 \pm 0,12$	$0,63 \pm 0,12$
WCN	$0,61 \pm 0,15$	$0,62 \pm 0,15$	$0,62 \pm 0,12$	$0,63 \pm 0,12$
AWCN	$0,59 \pm 0,16$	$0,60 \pm 0,16$	$0,60 \pm 0,13$	$0,62 \pm 0,13$

experimental representado por

$$\mathbf{b} = (b_1, b_2, \dots, b_N), \quad (5.3)$$

onde  $b_i$  é o fator-B experimental do  $i$ -ésimo  $C_\alpha$  da estrutura nativa obtida no PDB.

Nesta abordagem não é necessário diagonalizar a matriz Hessiana. As flutuações térmicas são calculadas diretamente da equação (5.1). A eliminação dessa etapa resulta em uma diminuição significativa de operações de ponto flutuante, a qual reduz o custo computacional.

Fazendo os cálculos novamente para as 125 proteínas em cada classe, mas agora utilizando este modelo, obtemos os resultados apresentados na tabela 3.

Na tabela 3, observa-se que os resultados fornecidos pelo AWCN são melhores que os do ANM. Além disto, ressaltamos que o tempo de processamento relativo a esse cálculo foi em média de 10 segundos enquanto que o tempo de processamento para o modelo ANM foi em média de 2 horas.

## 5.2 Expoente generalizado no potencial de interação entre resíduos

Nesta seção, iremos generalizar os modelos de rede elástica descritos anteriormente. Para isso, vamos escrever a matriz de Kirchhoff em função de um parâmetro  $p$ ,

$$\Gamma_{ij}^{pf} = \begin{cases} R_{ij}^{-p} & \text{se } i \neq j \\ -\sum_{i,j \neq i} \Gamma_{ij}, & \text{se } i = j \end{cases}$$

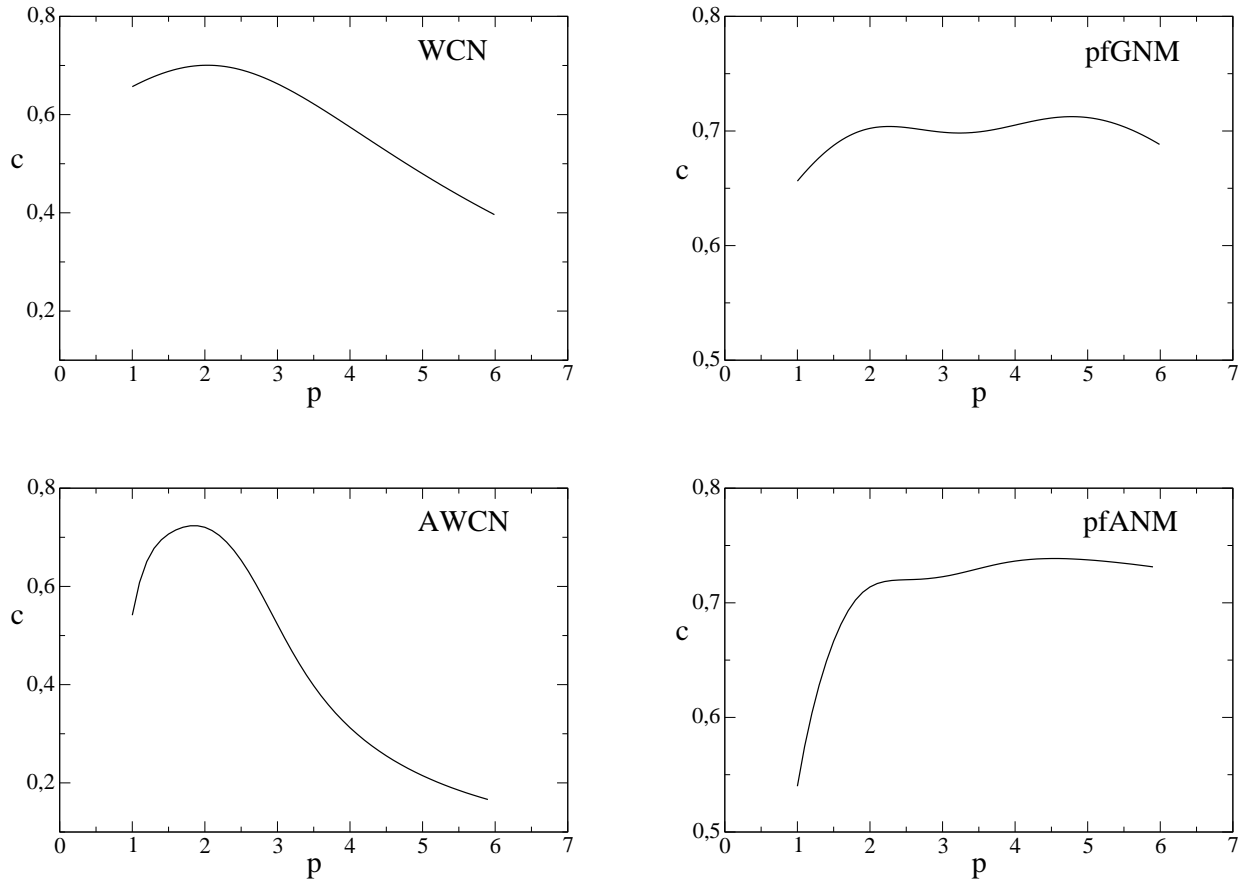


Figura 11: Comportamento do coeficiente de correlação  $c$  em função do parâmetro  $p$ , determinado pelos modelos WCN, pfGNM, pfANM e AWCN. Estas curvas estão associadas à proteína cujo código PDB é 16PK, a qual contém 425 resíduos.

e

$$H_{ij}^{pf} = H_{ij} R_{ij}^{-p}.$$

Nesta generalização, o caso  $p = 1$  corresponde aos modelos GNM e ANM. Enquanto que  $p = 2$  corresponde aos modelos de rede elástica livre de parâmetros pfGNM e pfANM. Entretanto, os potenciais clássicos descrevendo as interações entre átomos, aqui incluindo interações dipolares, quadripolares, são funções do inverso de  $r$ ,  $r^3$ ,  $r^6$ , ou seja  $r^p$ , sendo  $p$  um parâmetro empírico. Assim, para estudar o comportamento do coeficiente de correlação  $c$  em função do parâmetro  $p$  na equação acima, variamos  $p$  de 1 a 6, com passo de 0.01. Para cada  $p$ , calculamos o valor do  $c$  associado. Conforme ilustrado na figura 11, observa-se que para valores de  $p$  em torno de 2 para os modelos WCN e AWCN, o valor de  $c$  exibe o seu valor máximo. Para os modelos pfGNM e pfANM, observa-se que  $c$  assumi um valor máximo em torno de 2 e outro em torno de 4,75.

## 6 Conclusões e perspectivas

Apresentamos no capítulo 2 uma revisão sobre os modelos de rede elástica introduzidos desde a década de 80 (GÖ; NOGUTI; NISHIKAWA, 1993; TIRION, 1996; HALILOGLU; BAHAR; ERMAN, 1997; YANG; SONG; JERNIGAM, 2009). Mostramos, nesta revisão, que estes modelos evoluíram ao longo dos anos para modelos minimalistas. Nos primeiros modelos utilizavam-se potenciais complexos, enquanto que nos modelos atuais, como o GNM e ANM, as interações entre os resíduos são descritas por potenciais harmônicos.

Desenvolvemos programas computacionais de análise de modos normais para os modelos GNM, ANM, pfGNM, pfANM e WCN, com os quais comprovamos que os modelos pfGNM e pfANM descrevem melhor os dados experimentais que os modelos GNM e ANM tradicionais. Além disso, mostramos que para as 500 proteínas analisadas, embora todos os modelos sejam equivalentes, os modelos pfGNM, pfANM e WCN fornecem resultados superiores aos outros.

Propomos a versão anisotrópica do WCN intitulada AWCN. Mostramos que os resultados obtidos por meio dela apresentam melhor correlação que o ANM, além do ganho significativo de eficiência computacional.

Realizamos também um estudo de caráter investigativo do comportamento do expoente que pondera as interações entre os resíduos. Este estudo revelou que para os modelos WCN e AWCN a correlação exhibe o seu valor máximo para interações ponderadas pelo recíproco de  $R_{ij}^p$  para  $p \approx 2$ . Nos modelos pfGNM e pfANM a correlação é maximizada para dois valores de  $p$ , o primeiro em torno de 2 e o segundo em torno de 4,75, indicando que a ponderação pelo recíproco do quadrado da distância, usualmente empregada na literatura, pode não ser adequada para obter a melhor correlação.

Diversas proteínas apresentam estados que são classificados como, por exemplo, “aberto” ou “fechado”. Estes estados apresentam equilíbrio termodinâmico. A preferência entre estes estados é ditada por algum aspecto externo, possivelmente, por ligação com alguma molécula. É difícil seguir estas transições conformacionais, pois são, em geral, eventos que

acontecem em certa escala de tempo, exigindo destreza experimental. Mesmo simulações computacionais apresentam aspectos práticos de difícil contorno pois tais movimentos estão relacionados à observação de oscilações de grandes magnitudes na escala atômica e de baixas frequências. Por este motivo, modelos simplificados que descrevem a dinâmica vibracional destas proteínas apresentam a possibilidade de fornecer informações que de outra maneira, ou seria extremamente trabalhosa do ponto de vista experimental ou necessitaria de simulações computacionais que tornariam lento este estudo. Exemplos de proteínas que apresentam transições entre conformações termodinamicamente estáveis são as carregadoras de níquel e inclusive as que conduzem a doenças classificadas em famílias tais como Tauopatias, Poly-Q e de Alzheimer. Embora estas patologias sejam distintas, todas parecem compartilhar o mesmo mecanismo: todas envolvem a formação de agregados e degeneração de um conjunto específico de células neurais. A hipótese que prevalece para a explicação da agregação envolve uma mudança estrutural da cadeia polipeptídica de uma conformação nativa para uma inapropriadamente enovelada. Grande esforço tem sido dedicado para a compreensão do mecanismo molecular que provoca esta transição. Assim, acreditamos que este projeto de pesquisa deva focar futuramente a dinâmica intrínseca associada a cada conformação por meio dos modos lentos de vibração e caracterizar assim aspectos de ligação entre proteínas.

## Referências

- ALVES, N. A.; MARTINEZ, A. S. Inferring topological features of proteins from amino acid residue networks. *Physica A*, v. 375, p. 336–344, 2007.
- ATILGAN, A. R.; DURELL, S. T.; JERNIGAN, R. L.; DEMIREL, M. C.; KESKIN, O.; I.BAHAR. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal*, v. 80, p. 505–515, 2001.
- BAHAR, I.; ATILGAN; DURELL, S. T.; JERNIGAN, R. L.; DEMIREL, M. C.; KESKIN, O.; I.BAHAR. Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. *J. Mol. Biol.*, v. 80, p. 1023–1037, 1999.
- CUI, Q.; LI, G.; MA, J.; KARPLUS, M. A normal mode analysis of structural plasticity in the biomolecular motor F1-ATPase. *J. Mol. Biol.*, v. 340, p. 345–372, 2004.
- FLORY, P. J. Statistical thermodynamics of random networks. *Proc. R. Soc. Lond. A*, v. 351, p. 351–380, 1976.
- GŌ, N.; NOGUTI, T.; NISHIKAWA, T. Dynamics of a small globular protein in terms of low-frequency. *Proc. Natl. Acad. Sci. USA*, v. 80, p. 3696–3700, 1993.
- GOLDSTEIN, H. *Classical Mechanics*. Addison-Wesley: Chapman & Hall, 1980.
- GREENE, L. H.; HIGMAN, V. Uncovering network systems within protein structures. *J. Mol. Biol.*, v. 334, p. 781–791, 2003.
- HALILOGLU, T.; BAHAR, I.; ERMAN, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.*, v. 79, p. 3090–3093, 1997.
- HENZLER-WILDMAN, K.; KERN, D. Dynamic personalities of proteins. *NATURE*, v. 450, p. 964–972, 2007.
- KESKIN, O.; BAHAR, I.; FLATOW, D.; COVELL, D.; JERNIGAN, R. Molecular mechanisms of chaperonin of chaperonin GroEL-GroES function. *Biochemistry*, v. 41, p. 491–501, 2002.
- LIN, C.; SW, H.; YL, L.; SC, Y.; CH, S.; CH, L.; CC, H.; JK, H. Deriving protein dynamical properties from weighted protein contact number. *Proteins*, v. 72, p. 929–935, 2008.
- MARQUES, O. “BLZPACK: Description and User’s Guide”. TR/PA/95/30.CERFACS: France, 1995.
- MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamic of folded proteins. *NATURE*, v. 267, p. 585–590, 1997.

- MING, D.; KONG, Y.; WU, Y.; MA, J. Simulation of F-actin filaments of several microns. *Biophys. J*, v. 85, p. 27–35, 2003.
- MUWAD, L.; PERAHIA, D. Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures. *J. Mol. Biol.*, v. 258, p. 392–410, 1996.
- ORENGO, C. A.; THORNTON, J. Protein families and their evolution - a structural perspective. *Annu. Rev. Biochem.*, v. 74, p. 867–900, 2005.
- PRESS, W. H.; FLANNERY, B. P.; TEUKOLSKY, S. A.; VETTERLING, W. T. *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*. Hardcover: Cambridge, 1992.
- RADER, A. J.; VLAD, D. H.; BAHAR, I. Maturation dynamics of bacteriophage HK97 capsid. *Structure*, v. 13, p. 413–421, 2005.
- SHAKHNOVICH, E. I. Theoretical studies of proteins-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.*, v. 7, p. 29–40, 1997.
- TAMA, F.; VALLE, M.; FRANK, J.; BROOKS, C. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl. Acad. Sci. USA*, v. 100, p. 9319–9323, 2003.
- TEMIS, N. A.; BAHAR, I. V. Inhibitor binding alters the directions of domains motions in HIV-1 reverse transcriptase. *Proteins*, v. 49, p. 61–70, 2002.
- THOMAS, A.; HINSEN, K.; FIELD, M.; PERAHIA, D. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. *Proteins*, v. 34, p. 96–112, 1999.
- TIRION, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, v. 77, p. 1905–1908, 1996.
- WANG, Y.; RADER, A.; BAHAR, I.; JERNIGAN, R. L. Global ribosome motions revealed with elastic network model. *J. Struct. Biol.*, v. 147, p. 302–314, 2004.
- WILSON, E. B.; DECIUS, J.; CROSS, P. *Molecular Vibrations*. McGraw-Hill: New York, 1955.
- WYNSBERGHE, A. V.; LI, G.; CUI, Q. Normal-mode analysis suggests protein flexibility modulation throughout rna polymerase' s functional cycle. *Biochemistry*, v. 43, p. 13083–13095, 2004.
- YANG, L.; SONG, G.; JERNIGAN, R. L. Protein elastic network models and the ranges of cooperativity. *PNAS*, v. 106, p. 12347–12352, 2009.
- YANG, L.; SONG, G.; JERNIGAN, R. L. Comparisons of experimental and computed protein anisotropic temperature factors. *Proteins*, v. 76, p. 164–175, 2009.



## APÊNDICE A – Análise de modos normais (NMA)

Apresentamos neste apêndice a técnica intitulada: Análise de modos normais. Esta técnica baseia-se na aplicação da teoria clássica dos modos normais (GOLDSTEIN, 1980) na modelagem das oscilações ou flutuações térmicas dos átomos nas proteínas. A essência da técnica consiste na aproximação do potencial da proteína por um potencial quadrático (potencial harmônico de Hooke) com equação analítica simples e, em geral, envolvendo apenas um ou poucos parâmetros (TIRION, 1996).

### A.1 Teoria clássica dos modos normais

A proteína no estado nativo pode ser representada como um sistema com  $n$  graus de liberdade descritos pelas coordenadas generalizadas  $q_1, q_2, \dots, q_n$ . Estas coordenadas podem ser as coordenadas cartesianas, os ângulos entre as ligações, os ângulos torcionais, as coordenadas do centro de massa ou ainda qualquer conjunto de variáveis dinâmicas adequadas para descrever os graus de liberdade do sistema.

Para deslocamentos pequenos em torno de uma dada conformação de equilíbrio, representada pelo índice 0, a energia potencial  $V(\mathbf{q})$  da proteína pode ser expandida em série de Taylor,

$$V(\mathbf{q}) = V(\mathbf{q}^0) + \sum_{i=1}^n \left( \frac{\partial V}{\partial q_i} \right)_0 (q_i - q_i^0) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 (q_i - q_i^0)(q_j - q_j^0) + \dots$$

O primeiro termo da energia potencial está associado à conformação de equilíbrio. Escolhendo esta conformação como referência, este termo pode ser eliminado. Como na conformação de equilíbrio todos os átomos estão no ponto de mínimo da energia potencial, isto implica que

$$\left( \frac{\partial V}{\partial q_i} \right)_0 = 0, \quad i = 1, 2, \dots, n.$$

Desta forma, o segundo termo também pode ser eliminado.

Para deslocamentos suficientemente pequenos em torno das posições de equilíbrio, os termos de ordem maior do que dois podem ser desprezados. Assim a energia potencial se reduz à forma quadrática,

$$V(\mathbf{q}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n F_{ij} (q_i - q_i^0)(q_j - q_j^0) = \frac{1}{2} \mathbf{q}^T \mathbf{F} \mathbf{q}, \quad (\text{A.1})$$

onde  $F_{ij}$  são os elementos da matriz Hessiana  $\mathbf{F}$ , os quais são dados por

$$F_{ij} = \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0, \quad (\text{A.2})$$

com  $F_{ij} = F_{ji}$ . Analogamente, a energia cinética também pode ser expressa na forma quadrática nas velocidades generalizadas  $\dot{q}_i$ ,

$$E_c = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n H_{ij} \dot{q}_i \dot{q}_j = \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{H} \dot{\mathbf{q}}, \quad (\text{A.3})$$

onde os elementos da matriz  $\mathbf{H}$  são dados por

$$H_{ij} = \sum_{l=1}^N m_l \frac{\partial \mathbf{r}_l}{\partial q_i} \frac{\partial \mathbf{r}_l}{\partial q_j} \quad (\text{A.4})$$

(ver apêndice B.1). As equações do movimento de Newton para os átomos  $i$  são obtidas por meio das equações de Lagrange,

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) = \frac{\partial L}{\partial q_i}, \quad (\text{A.5})$$

onde a Lagrangiana  $L$  é dada por

$$L = E_c - E_p. \quad (\text{A.6})$$

Derivando parcialmente as equações (A.1) e (A.3), respectivamente com relação a  $q$  e a  $\dot{q}$ , obtemos as equações:

$$\frac{\partial L}{\partial \dot{q}_i} = \sum_{j=1}^n H_{ij} \dot{q}_j \quad \text{e} \quad \frac{\partial L}{\partial q} = - \sum_{j=1}^n F_{ij} (q_j - q_j^0).$$

Substituindo estas equações na equação (A.5), obtemos o sistema de equações do movimento

$$\sum_{j=1}^n H_{ij} \ddot{q}_j = - \sum_{j=1}^n F_{ij} (q_j - q_j^0). \quad (\text{A.7})$$

Este sistema de equações diferenciais acopladas é resolvido assumindo que a solução

é o vetor  $\mathbf{q}$  com coordenadas  $q_j$  da forma:

$$q_j = q_j^0 + \sum_{k=1}^n A_{jk} \alpha_k \cos(\omega_k t + \delta_k). \quad (\text{A.8})$$

Ou seja, as vibrações do sistema podem ser expressas como uma superposição de osciladores independentes (modos normais).

Da equação (A.8), facilmente obtemos

$$\ddot{q}_j = \frac{d^2 q_j}{dt^2} = - \sum_{k=1}^n \omega_k^2 A_{jk} \alpha_k \cos(\omega_k t + \delta_k).$$

Substituindo  $q_j$  e  $\ddot{q}_j$  na equação (A.7) resulta em

$$- \sum_{i=1}^n H_{ij} \sum_{k=1}^n \omega_k^2 A_{jk} \alpha_k \cos(\omega_k t + \delta_k) = \sum_{j=1}^n F_{ij} \sum_{k=1}^n A_{jk} \alpha_k \cos(\omega_k t + \delta_k).$$

Esta equação pode ser reescrita como

$$- \sum_k \omega_k^2 \left( \sum_{j=1}^n H_{ij} A_{jk} \right) \alpha_k \cos(\omega_k t + \delta_k) = \sum_{k=1}^n \left( \sum_{j=1}^n F_{ij} A_{jk} \right) \alpha_k \cos(\omega_k t + \delta_k). \quad (\text{A.9})$$

A identidade na equação (A.9) é válida se e somente se

$$\omega_k^2 \left( \sum_{j=1}^n H_{ij} A_{jk} \right) = \sum_{j=1}^n F_{ij} A_{jk}, \quad (\text{A.10})$$

para  $k = 1, \dots, n$ .

Reescrevendo o sistema de equações representado na equação (A.10) na notação matricial, obtemos a equação do problema de autovalores,

$$\mathbf{FA} = \mathbf{\Lambda HA}, \quad (\text{A.11})$$

sujeita a condição de normalização  $\mathbf{A}^T \mathbf{HA} = \mathbf{I}$ , a qual garante que a Hamiltoniana é diagonalizada (B.2). Na equação (A.11)  $\mathbf{A}$  é a matriz cujas colunas são os autovetores da matriz  $\mathbf{F}$ ,  $\mathbf{\Lambda}$  é a matriz diagonal dos autovalores com  $\Lambda_{ii} = \omega_i^2$  e  $\omega_i$  são as frequências normais de vibração do sistema. Em suma, as frequências normais e os autovetores podem ser determinados diagonalizando a matriz Hessiana  $\mathbf{F}$ . No entanto, para esta matriz ser determinada é necessário o conhecimento da configuração de equilíbrio a priori, a qual pode ser obtida por meio da minimização da energia potencial.

## APÊNDICE B – Demonstrações

### B.1 Demonstração do cálculo de $H_{ij}$

A energia cinética de um sistema de  $N$  partículas com massa  $m_l$  em coordenadas cartesianas é dada por

$$E_c = \frac{1}{2} \sum_{l=1}^N m_l \dot{\mathbf{r}}_l^2. \quad (\text{B.1})$$

Variações pequenas nas coordenadas cartesianas  $\delta \mathbf{r}_l$  podem ser expressas em função das coordenadas generalizadas da seguinte forma

$$\delta \mathbf{r}_l = \sum_{i=1}^n \left( \frac{\partial \mathbf{r}_l}{\partial q_i} \right) \delta q_i,$$

Considerando que estas variações ocorrem em um intervalo de tempo  $\delta t$ , podemos escrever

$$\delta \mathbf{r}_l / \delta t = \sum_{i=1}^n \left( \frac{\partial \mathbf{r}_l}{\partial q_i} \right) \delta q_i / \delta t.$$

substituindo esta equação na equação (B.1) resulta:

$$\begin{aligned} E_k &= \frac{1}{2} \sum_{l=1}^N m_l \sum_{i=1}^n \frac{\partial \mathbf{r}_l}{\partial q_i} \dot{q}_i \sum_{j=1}^n \frac{\partial \mathbf{r}_l}{\partial q_j} \dot{q}_j \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left( \sum_{l=1}^N m_l \frac{\partial \mathbf{r}_l}{\partial q_i} \frac{\partial \mathbf{r}_l}{\partial q_j} \right) \dot{q}_i \dot{q}_j, \end{aligned}$$

comparando com a equação da energia cinética (A.3) temos os elementos de matriz

$$H_{ij} = \sum_{l=1}^N m_l \frac{\partial \mathbf{r}_l}{\partial q_i} \frac{\partial \mathbf{r}_l}{\partial q_j}. \quad (\text{B.2})$$

## B.2 Condição de normalização

Antes de resolvermos a equação (A.11) é necessário definirmos a condição de normalização. Definindo  $Q_k = \alpha_k \cos(\omega_k t + \delta_k)$ , a equação (A.8) pode ser reescrita como

$$q_j - q_j^0 = \sum_{k=1}^n A_{jk} Q_k = \mathbf{A}\mathbf{Q}. \quad (\text{B.3})$$

A equação (B.3) expressa uma relação, transformação linear, entre as coordenadas generalizadas  $\mathbf{q}$  e as coordenadas normais  $\mathbf{Q}$ .

Fisicamente, os modos normais de vibração do sistema se comportam como um conjunto (*ensemble*) de osciladores harmônicos independentes. Assim, a Hamiltoniana do sistema escrita em função das coordenadas normais  $Q_k$  deverá possuir uma forma diagonal. Escrevendo a energia cinética e a potencial em termos de  $Q$  resulta em:

$$E_p = \frac{1}{2}(\mathbf{q} - \mathbf{q}^0)^T \mathbf{F}(\mathbf{q} - \mathbf{q}^0) = \frac{1}{2} \mathbf{Q}^T \mathbf{A}^T \mathbf{F} \mathbf{A} \mathbf{Q} \quad (\text{B.4})$$

$$E_c = \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{H} \dot{\mathbf{q}} = \frac{1}{2} \dot{\mathbf{Q}}^T \mathbf{A}^T \mathbf{H} \mathbf{A} \dot{\mathbf{Q}}. \quad (\text{B.5})$$

Se impormos a condição de normalização

$$\mathbf{A}^T \mathbf{H} \mathbf{A} = \mathbf{1},$$

a energia cinética assume a forma

$$E_c = \frac{1}{2} \mathbf{Q}^T \dot{\mathbf{Q}} = \frac{1}{2} \sum_{k=1}^n \dot{Q}_k^2. \quad (\text{B.6})$$

Esta condição implica também que se multiplicarmos a equação (A.11) pela esquerda por  $\mathbf{A}^T$ ,

$$\mathbf{A}^T \mathbf{F} \mathbf{A} = \mathbf{A}^T \mathbf{H} \mathbf{A} \mathbf{\Lambda} = \mathbf{\Lambda}.$$

Assim, a energia potencial pode ser expressa como a soma das energias de osciladores desacoplados com frequências  $\omega_k$  e deslocamentos  $Q_k$ ,

$$E_p = \frac{1}{2} \mathbf{Q}^T \mathbf{\Lambda} \mathbf{Q} = \frac{1}{2} \sum_{k=1}^n \Lambda_{kk} Q_k^2 = \frac{1}{2} \sum_{k=1}^n \omega_k^2 Q_k^2. \quad (\text{B.7})$$

Pelo fato de que as energias cinética e potencial são expressas como simples somas das coordenadas  $Q_k$  e  $\dot{Q}_k$ , a Hamiltoniana do sistema  $\mathcal{H} = E_c + E_p$  assume a forma diagonal. Portanto, a matriz  $\mathbf{A}$  diagonaliza a Hamiltoniana do sistema.

Como a energia cinética

$$T = \frac{1}{2} \sum_{k=1}^n \dot{Q}_k^2$$

é escrita como a soma das coordenadas  $\dot{Q}$ , as coordenadas  $Q$  são ponderadas pelas massas.

O comportamento dinâmico do sistema é completamente descrito pela equação

$$q_j = q_j^0 + \sum_{k=1}^n A_{jk} \alpha_k \cos(\omega_k t + \delta_k).$$

## APÊNDICE C – Teoria de grafos

### C.1 Uma breve introdução

Basicamente, um grafo é definido com sendo um par  $G = (V, E)$  com conjunto de vértices  $V = \{v_1, v_2, \dots, v_n\}$  e um conjunto de arestas  $E = E(G) = \{e_1, e_2, \dots, e_n\}$ . Cada aresta é definida como  $e_i = \{v_i, v_j\}$ . Se escolhermos umas das extremidades  $i$  e  $j$  como sendo o sentido positivo ou negativo, dizemos que o grafo é orientado. Nesta dissertação, trabalharemos com grafos não orientados.

A matriz *adjacência*  $A = (a_{ij})$  é definida por

$$a_{ij} := \text{multiplicidade de arestas } (v_i, v_j) \text{ em } E$$

i.e.,  $a_{ij}$  é igual ao número de arcos que sai de um vértice  $v_i$  e incide no vértice  $v_j$ . O espectro de  $G$  é o espectro da matriz adjacência.

#### C.1.1 A matriz Laplaciana

A título de exemplo, resolveremos a seguir o problema clássico das vibrações de membranas para justificarmos historicamente a origem do nome matriz Laplaciana.

Considere uma membrana no plano  $xy$ , com deslocamento vertical  $z = z(x, y)$ , a equação de onda é dada por

$$\nabla^2 z(x, y) = \frac{1}{c^2} \frac{\partial^2 z}{\partial t^2}, \quad (\text{C.1})$$

sendo  $t$  a variável tempo e  $c$  a velocidade da onda no meio. Por outro lado,  $\nabla^2 z(x, y) = \partial^2 z / \partial x^2 + \partial^2 z / \partial y^2$ . Se assumirmos que a membrana se comporta como uma mola, existirá uma força restauradora proporcional ao deslocamento dada pela lei de Hooke,

$$m \frac{\partial^2 z}{\partial t^2} = -kz \quad \text{ou seja} \quad \frac{\partial^2 z}{\partial t^2} = -Kz,$$

em que  $K = k/m$ .

Aproximando a membrana por uma grade discreta de partículas de espaçamento  $h$ , as derivadas parciais podem ser aproximadas por uma função  $z(x)$ , a qual pode ser expandida em série de Taylor como

$$z(x) = z(x_0) + \frac{dz(x)}{dx} \frac{(x - x_0)}{1!} + \frac{d^2z(x)}{dx^2} \frac{(x - x_0)^2}{2!} + \dots$$

Fazendo  $x_0 = x$ ,  $x' = x+h$  e  $x'' = x-h$ , temos que  $z(x')$  e  $z(x'')$  podem ser respectivamente escritos como

$$z(x+h) = z(x) + \frac{dz(x)}{dx}h + \frac{1}{2} \frac{d^2z(x)}{dx^2}h^2 + \dots$$

$$z(x-h) = z(x) - \frac{dz(x)}{dx}h + \frac{1}{2} \frac{d^2z(x)}{dx^2}h^2 + \dots$$

Somando as duas últimas equações e desprezando os termos de ordem superiores obtemos

$$z(x+h) + z(x-h) \approx 2z(x) + \frac{d^2z(x)}{dx^2}h^2 \Rightarrow \frac{d^2z(x)}{dx^2} \approx \frac{z(x+h) + z(x-h) - 2z(x)}{h^2}.$$

Portanto,

$$\frac{\partial^2 z}{\partial x^2}(x, y) \approx \frac{z(x+h, y) + z(x-h, y) - 2z(x, y)}{h^2} \quad (\text{C.2})$$

$$\frac{\partial^2 z}{\partial x^2}(x, y) \approx \frac{z(x+h, y) + z(x-h, y) - 2z(x, y)}{h^2} \quad (\text{C.3})$$

$$\frac{\partial^2 z}{\partial y^2}(x, y) \approx \frac{z(x, y+h) + z(x, y-h) - 2z(x, y)}{h^2} \quad (\text{C.4})$$

Substituindo as equações (C.3) e (C.4) na equação (C.1) obtemos

$$4z(x, y) - z(x+h, y) - z(x, y+h) - z(x, y-h) \approx \frac{Kh^2}{c^2}z(x, y)$$

Isto diz que  $z$  é uma autofunção da matriz Laplaciana da grade do grafo com autovalores  $kh^2/c^2$ . Isto explica o termo Laplaciana para matriz  $L$  porque suas funções são análogas discretas do operador Laplaciano contínuo.



## C.2 Matriz de Kirchhoff

Considere  $G = (V, E)$  sendo um multigrafo não direcionado com matriz adjacência  $A$ . Seja  $D = \text{diag}(d(1), \dots, d(n))$  sendo a matriz diagonal dos graus dos vértices. A matriz Laplaciana  $\Gamma = (\gamma_{ij})$  é definida como  $L := D - A$ , assim se  $G$  é um grafo não direcionado simples, então

$$\Gamma_{ij} = \begin{cases} -1 & \text{se } \{i, j\} \in E \\ d(i) & \text{se } i = j \\ 0 & \text{caso contrário.} \end{cases}$$

## C.3 Cálculo dos elementos da matriz Hessiana no ANM

Considere dois osciladores, representados pelos carbonos  $\alpha$   $i$  e  $j$ , conectados por uma mola virtual com constante de força  $\gamma$ , a energia potencial deste sistema pode ser escrita como

$$V = \frac{1}{2}\gamma(d_{ij} - d_{ij}^0)^2, \quad (\text{C.5})$$

onde  $d_{ij} = [(X_j - X_i)^2 + (Y_j - Y_i)^2 + (Z_j - Z_i)^2]^{1/2}$  é o comprimento da mola distendida e  $d_{ij}^0$  é o seu comprimento natural. Fazendo a mudança de variável  $u(X_i, Y_i, Z_i) = d_{ij}(X_i, Y_i, Z_i) - d_{ij}^0(X_i, Y_i, Z_i)$ , pela regra da cadeia podemos escrever

$$\frac{\partial V}{\partial X_i} = \frac{\partial V}{\partial u} \frac{\partial u}{\partial X_i} \quad (\text{C.6})$$

$$\begin{aligned} &= \gamma u \frac{\partial u}{\partial X_i} \\ &= -\gamma(d_{ij} - d_{ij}^0) \frac{(X_j - X_i)}{d_{ij}} \\ &= -\gamma \left(1 - \frac{d_{ij}^0}{d_{ij}}\right) (X_j - X_i) \end{aligned} \quad (\text{C.7})$$

Fazendo  $D = -\gamma \left(1 - \frac{d_{ij}^0}{d_{ij}}\right)$  e  $X = (X_j - X_i)$  na equação (C.7), a derivada segunda da energia potencial com relação a coordenada  $X_i$  pode ser calculada pela regra da derivada do produto,

$$\frac{\partial^2 V}{\partial X_i \partial X_i} = \frac{\partial D}{\partial X_i} X + D \frac{\partial X}{\partial X_i}. \quad (\text{C.8})$$

O termo  $\frac{\partial D}{\partial X_i}$  pode ser calculado da seguinte forma:

$$\frac{\partial D}{\partial X_i} = \gamma d_{ij}^0 \frac{\partial(d_{ij}^{-1})}{\partial X_i} \quad (\text{C.9})$$

$$\begin{aligned} &= -\gamma d_{ij}^0 d_{ij}^{-2} \frac{\partial(d_{ij})}{\partial X_i} \\ &= \gamma d_{ij}^0 d_{ij}^{-3} (X_j - X_i). \end{aligned} \quad (\text{C.10})$$

Substituindo a equação (C.10) na equação (C.8), obtemos

$$\begin{aligned} \frac{\partial^2 V}{\partial X_i \partial X_i} &= \gamma d_{ij}^0 d_{ij}^{-3} (X_j - X_i)(X_j - X_i) + \gamma \left(1 - \frac{d_{ij}^0}{d_{ij}}\right) \\ &= \gamma \left(1 + d_{ij}^0 \frac{(X_j - X_i)^2}{d_{ij}^3} - \frac{d_{ij}^0}{d_{ij}}\right). \end{aligned} \quad (\text{C.11})$$

Assim, a derivada segunda da energia potencial com relação a coordenada  $X_i$ , calculada na conformação de equilíbrio, é dada por

$$\left(\frac{\partial^2 V}{\partial X_i \partial X_i}\right)_0 = \gamma \frac{(X_j - X_i)^2}{d_{ij}^{0,2}}. \quad (\text{C.12})$$

Analogamente,

$$\left(\frac{\partial^2 V}{\partial Y_i \partial Y_i}\right)_0 = \gamma \frac{(Y_j - Y_i)^2}{d_{ij}^{0,2}} \quad \text{e} \quad \left(\frac{\partial^2 V}{\partial Z_i \partial Z_i}\right)_0 = \gamma \frac{(Z_j - Z_i)^2}{d_{ij}^{0,2}}. \quad (\text{C.13})$$

No caso em que o  $i$ -ésimo  $C_\alpha$  está conectado a  $\Gamma_{ii}$  vizinhos, a equação (C.12) pode ser reescrita como:

$$\left(\frac{\partial^2 V}{\partial X_i \partial X_i}\right)_0 = \gamma \sum_j \frac{(X_j - X_i)^2}{d_{ij}^{0,2}}, \quad (\text{C.14})$$

onde a soma é realizada sobre todos os  $\Gamma_{ii}$   $C_\alpha$ , ou seja, para  $j = 1, \Gamma_{ii}$ . Derivando a equação (C.7) com relação a  $Y_j$ , obtemos

$$\begin{aligned} \frac{\partial^2 V}{\partial Y_j \partial X_i} &= \frac{\partial D}{\partial Y_j} X + D \frac{\partial X}{\partial Y_j} \\ &= -\gamma \frac{(Y_j - Y_i)(X_j - X_i)}{d_{ij}^{0,2}}, \end{aligned} \quad (\text{C.15})$$

$$\mathbf{H}_{ij} = \begin{bmatrix} \frac{\partial^2 V}{\partial X_i \partial X_j} & \frac{\partial^2 V}{\partial X_i \partial Y_j} & \frac{\partial^2 V}{\partial X_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Y_i \partial X_j} & \frac{\partial^2 V}{\partial Y_i \partial Y_j} & \frac{\partial^2 V}{\partial Y_i \partial Z_j} \\ \frac{\partial^2 V}{\partial Z_i \partial X_j} & \frac{\partial^2 V}{\partial Z_i \partial Y_j} & \frac{\partial^2 V}{\partial Z_i \partial Z_j} \end{bmatrix} \quad (\text{C.16})$$

e

$$\mathbf{H}_{ij} = \frac{-\gamma}{d_{ij}^0} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}. \quad (\text{C.17})$$

## APÊNDICE D – Programas

### D.1 GNM

```

program gaussian_network_model
cc %-----%
cc | Este programa calcula os coeficientes de correlação de person entre |
cc | o fator-B teorico/experimental para np proteínas por meio do modelo |
cc | de rede gaussiana (GNM) |
cc | |
cc %----- referencia----- %
cc | HALILOGLU, T.; BAHAR, I.; ERMAN, B. Gaussian dynamics of folded |
cc | proteins. Phys. Rev. Lett., v. 79, n. 16, p. 3090{3093, Outubro 1997. |
cc % -----%
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc %-----input-----%
cc | np: # de proteínas |
cc | exemplo de arquivo de entrada=> 1A6M.txt o qual contem as coordenadas |
cc | |x,y,z e o fator-B de temperatura dos carbonos-alpha. |
cc %-----%
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc Kirc: matriz de kirchhoff (n,n), n: # de carbonos alpha
cc aut: vetor de ordem n contendo n autovalores da matriz de Kirchhoff
cc vec: matriz dos autovetores vec(n,n)
cc invgm: matriz pseudoinversa (inversa generalizada)da matriz de kirchhoff
cc cperson:coeficiente de correlacao de person entre fator-B teorico
cc e experimental
cc bf: vetor cujas componentes são os B-factors experimentais
cc inv: vetor onde cada componente é o recíproco do seu autovalor associado

```

```

cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
parameter(np=125)
integer*4 i,j,n,kk !kk: # de experimentos
real*8,dimension(:),allocatable :: aut,corre,x,y,z,inv,cperson,bf
real*8 soma,dist
real*8,dimension(:,:),allocatable :: invgm,vec, kirc
real*8 distx,disty,distz,r,pi
character*80 arquivo2,list,lista,arquivo3,nome2,nome
pi=3.14159265359d0
cc criar um novo arquivo para escrever os coeficientes de correlacao
      open(unit=70,file='cperson_gnm_all_beta.dat')
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc %-----%
cc |abrir uma tabela de strings e # a qual contem informacoes sobre o codigo |
cc |pdb da proteina, # de carbonos alpha e um label associado a ela. Exemplo |
cc |das quatro primeiras linhas da lista: |
cc |  1A6M      151      1 |
cc |  1AE7      119      2 |
cc |  1AF7      274      3 |
cc |  1AH7      245      4 |
cc %-----%
      lista='/home/users/pdb_data/list_pdb/list_99_beta.txt'!diretorio da lista
      l=iendst(lista)
      arquivo2=lista(1:1)
      open(4,file=arquivo2,status='old')
cc %-----producao-----%
cc |por exemplo, para kk=1,é feita a leitura do codigo (1A6M) da proteina |
cc |e do seu # de carbonos-alpha (151). Em seguida, é feita a leitura do |
cc |arquivo 1A6M.txt no diretorio onde ele se encontra. |
cc %-----%
      do kk=1,np
      read(4,*) nome,n
      allocate (aut(1:n),corre(1:n),inv(1:n),cperson(1:np),bf(1:n))
      allocate (x(1:n),y(1:n),z(1:n))
      allocate (invgm(1:n,1:n),vec(1:n,1:n),kirc(1:n,1:n))

```

```

mm=iendst(nome)
list=nome(1:mm)//'.txt' !para k=1, list=1A6M.txt
cc %-----%
cc | no diretório '/home/users/mendonca/eigenvalues/GNM/all_alpha_99/' utili- |
cc | zando o programa read_clean_pdb_to_ca.f foram criados arquivos.txt refe- |
cc | rentes as proteínas listadas acima. Cada arquivo contem as coordenadas |
cc | x,y,z e o fator-B de temperatura dos carbonos-alpha de uma dada proteína |
c | da lista. |
cc %-----%
nome2='/home/users/mendonca/eigenvalues/GNM/all_beta_99/'
ll=iendst(nome2)
arquivo3=nome2(1:ll)//list
print*,arquivo3
open(1,file=arquivo3,status='old')
do i=1,n
read(1,*) x(i),y(i),z(i),bf(i) !ler as coordenadas x,y,z e b-factor
end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c inicialização da matriz de kirchhoff
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc zerando todos os elementos
do i=1,n
do j=1,n
kirc(i,j)=0.d0
end do
end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc construindo os elementos fora da diagonal
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
do i=1,n-1
do j=i+1,n
distx=(x(i)-x(j))
disty=(y(i)-y(j))
distz=(z(i)-z(j))
dist=dsqrt(distx*distx+disty*disty+distz*distz)

```

```

        if(dist.le.7.3d0)then
            kirc(i,j)=-1.d0
            kirc(j,i)=-1.d0
        end if
    end do
end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc                    construindo os elementos da diagonal
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
do i=1,n
    soma=0.d0
    do j=1,n
        soma=soma+kirc(i,j)
    end do
    kirc(i,i)=-soma
end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc            Obtendo os autovalores e autovetores da matriz de kirchhoff
cc a subrotina svdcmp tambem pode ser utilizada
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c            call svdcmp(kirc,n,n,n,n,aut,vec)
            call jacobi(kirc,n,n,aut,vec,nrot)
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c            ordenando os autovalores e os respectivos autovetores em ordem crescente
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
            call eigsrt(aut,vec,n,n)
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc            calculo do reciproco dos autovalores
            do i=1,n
                inv(i)=1/aut(i)
            end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc                    construindo a matriz pseudoinversa
cc o primeiro autovalor (aut(1)=0.) correspondente a translacao do sistema como
cc um todo é eliminado

```

```
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
do i=1,n
  do j=1,n
    invgm(i,j)=0.d0
    do k=2,n
      invgm(i,j)=invgm(i,j)+inv(k)*vec(i,k)*vec(j,k)
    end do
  end do
end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc          construindo o vetor associado a  autocorrelacao das flutuacoes
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
do i=1,n
do j=1,n
corre(i)=invgm(i,i)
end do
end do
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc
cc  obtencao do coeficiente de correlacao de person entre os B-factor teorico
ec    e o experimental
cc  armazenamento do fator-B de cada proteina nas componentes do vetor
c person(np)
cc
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
  call pearsn(bf,corre,n,r)
  cperson(kk)=r
  write(70,*) n,cperson(kk) !escrevendo o label da proteina e a correlacao
cc21  format(F7.4)
cc    end do
cc  %-----%
cc  |           0 fator de temperatura teórico é dado por           |
cc  |           Bteorico=((8*pi*pi*kT)/gamma)*invgm(i,i)           |
cc  |           =co*corre(i)           |
cc  %-----%
```



```

cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc          normalizando os B-factors  teoricos com os experimentais
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc      call fit(corre,bf,co,n)
cc      print*, 'gamma:', (8*pi*pi)/co
cc      do i=1,n
cc      write(90,*)  i,co*corre(i)!  escrever o label da proteina e seu fator-B
cc      teorico
cc
cc      end do
cc      deallocate (aut,corre,inv,cperson,bf)
cc      deallocate (x,y,z)
cc      deallocate (invgm,vec,kirc)
cc      end do !uma realizacao
cc      stop
cc      end
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc          Subrotinas
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
cc          esta subrotina calcula o coeficiente de correlacao de person r entre
cc          dois vetores x(n) e y(n)
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
      SUBROUTINE pearsn(x,y,n,r)
      INTEGER*4 n
      REAL*8 r,x(n),y(n),TINY
      PARAMETER (TINY=1.d-20)
CU    USES betai
      INTEGER*4 j
      REAL*8 ax,ay,sxx,sxy,syy,xt,yt
      ax=0.d0
      ay=0.d0
      do 11 j=1,n
          ax=ax+x(j)
          ay=ay+y(j)
11    continue

```

```

ax=ax/n
ay=ay/n
sxx=0.d0
syy=0.d0
sxy=0.d0
do 12 j=1,n
    xt=x(j)-ax
    yt=y(j)-ay
    sxx=sxx+xt**2
    syy=syy+yt**2
    sxy=sxy+xt*yt
12  continue
    r=sxy/(dsqrt(sxx*syy)+TINY)

cc    z=0.5*dlog(((1.+r)+TINY)/((1.-r)+TINY))
cc    df=n-2
cc    t=r*dsqrt(df/(((1.-r)+TINY)*((1.+r)+TINY)))
cc    prob=betai(0.5*df,0.5,df/(df+t**2))
cc    prob=erfcc(dabs(z*dsqrt(n-1.))/1.4142136)
    return
    END

cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
c          normalizacao do b-teorico com o experimental
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

    subroutine fit(z,w,a,n)
    integer*4 n
    real*8 z(n),w(n),sb,sc,a
    sb=0.d0
    sc=0.d0
    do i=1,n
        sb=sb+w(i)
        sc=sc+z(i)
    end do
    a=sb/sc
    return

```





## APÊNDICE E – Lista das proteínas analisadas

Os códigos apresentados na tabela abaixo identificam as proteínas utilizadas nesta dissertação.

Tabela 4: Código PDB e o número de resíduos  $N$  para 125 proteínas em cada classe de acordo com a classificação SCOP.

all- $\alpha$	$N$	all- $\beta$	$N$	$\alpha+\beta$	$N$	$\alpha/\beta$	$N$
1JHG	101	1KDJ	102	1P1L	102	1GMX	107
1L3P	102	1FMB	104	1KCQ	103	1JF8	130
3CAO	102	1BTN	106	1NZA	103	1J7G	144
5CYT	103	1ACX	108	1ONC	103	1I5G	144
1TQG	105	1NPL	109	1I0V	104	1F4P	147
1GXQ	105	1NOA	113	1FUS	105	1DG9	157
1M6T	106	1NPU	116	1EW4	106	1DF7	159
1CO6	107	1L7L	121	1BKF	107	1HD2	161
1UP9	107	1BQK	124	1CEW	108	1CTQ	166
1YCC	107	1DFX	125	1OAP	108	1KAO	167
2PVB	107	1I3U	126	1I4M	108	1JWQ	179
1BKR	108	1IJT	128	1N9L	109	1IM5	179
1J0P	108	1KNM	129	1KM8	110	1EJ0	180
1RRO	108	1CUO	129	1E87	117	1BED	181
1S3P	109	1NEP	130	1GNU	117	1KMV	185
1KR7	110	1HMT	131	1BUO	121	1EX7	186
1WAD	111	1IFC	131	1DHN	121	1A2J	188
3C2C	112	1LMI	131	1QTO	122	1JUV	193
1I8O	113	1MDC	131	1C44	123	1IJB	202
1DLW	116	1LPJ	133	1HQ8	123	1DYR	205
2CY3	118	1IS3	134	1B9O	123	1HDO	205
1POA	118	1DQG	134	1KF3	124	1H2E	207
2MHR	118	1KQW	134	1NWZ	125	1G66	207
1AE7	119	1GMI	135	1AGI	125	1IO2	213

1LWB	122	1FSO	138	1ACF	125	1JFX	217
1MC2	122	1EJ8	140	1JUG	125	1B4P	217
1VPI	122	1IFG	140	1DOI	128	1AKY	218
1ZLB	122	1H9K	144	1MSC	129	1I1N	224
1G4I	123	1ALY	146	1EW0	130	1BYI	224
1S69	123	1NB9	147	1QQY	130	1A8L	226
1BGF	124	1OAL	151	1C7K	132	1G8A	227
1CXC	124	1JCV	153	1H1H	134	1G8S	230
1CPQ	129	1GWM	153	1GQV	135	1ATG	231
2B8U	129	1GNY	153	1PNE	139	1AMF	231
2NSZ	129	1MFM	153	1Q2Y	140	1K7C	233
1PA7	130	1MD6	154	1LO7	140	1G3Q	237
1C52	131	1ESO	154	1H6H	143	1LST	238
1POC	134	1GUI	155	1QDD	144	1LYX	246
1E29	135	1KEX	155	1ICX	155	1A53	247
1ECA	136	1K12	158	1BD8	156	1M2K	249
2END	137	1D7P	159	1G1T	157	1GQN	252
1B0B	141	1CZT	160	1Q0N	158	1B2L	254
1TZV	141	1J2A	166	1FM4	159	1A8P	257
1K6K	142	1D2S	170	1QST	160	1GS5	258
1EYH	144	1CCZ	171	1I76	163	1FSF	266
1MBA	146	1E3B	172	1CJW	166	1GCI	269
1MYT	146	1H4A	173	1G12	167	1E0C	270
1JMW	146	1I40	175	1CIZ	168	1A8S	273
2HBG	147	1A58	177	1KOE	172	1A8Q	274
1KFR	147	1CDY	178	1CV8	173	1AF7	274
1DK8	147	1CQQ	180	1EB6	177	1BRT	277
1NG6	148	1H0P	182	1LQY	184	1AJ2	282
2LHB	149	1A44	185	1GBS	185	1DQY	283
1A6M	151	1I8A	189	1JHS	188	1H6T	291
1SRA	151	1JM1	202	1BUD	197	1KS9	291
1ILK	151	1CFB	205	1OCY	198	1J8M	295
1DVO	152	1MIX	206	1IAE	200	1JLN	297
2GDM	153	1AUN	208	1IQQ	200	1D7O	297
1LHT	153	1ENF	212	1IAG	201	1BQC	302
2FD5	180	1LO6	221	1KUF	201	1I1W	302
1PI1	185	1HJ9	223	1ND1	201	1IOW	306
1LKP	190	1AOL	228	1DIX	208	1B6G	308
1OMR	201	1BIO	228	1QNX	209	1H6U	308
1MIX	206	1ELT	236	1COJ	211	1EA7	309
2O7O	207	1NLS	237	1ENF	212	1FL2	310
1COJ	211	1GVZ	237	1OJQ	212	1GSA	314
2ABK	211	1BEC	238	1PPN	212	1CVL	316
1FT5	211	1GZC	239	1LBU	213	1HNJ	317
1SDI	213	1EAX	241	1ME3	215	1IXH	321
1LBU	213	1LED	242	1IWD	215	1B4E	323
1C1K	217	1DUA	242	1PPO	216	1BG2	323
1B4P	217	1CI3	249	1BOL	222	1CWN	324
1MUN	225	1HCZ	250	1QR0	228	1B8P	327

1Y6I	233	1A8P	257	1KO3	230	1IEJ	329
1FCY	236	1HCB	258	1FX2	235	1DXY	330
1AH7	245	1FLJ	259	1DXJ	242	1J33	333
1EG3	260	1LUG	259	1QTS	247	1FOB	334
1AF7	274	1EG3	260	1MRJ	247	1BX4	342
1QQF	276	1JOV	269	1H70	255	1EK5	346
1KS9	291	1F00	282	1QGI	259	1EZW	347
1J8M	295	1H6T	291	1J1Q	261	1JIX	351
2CIW	298	1H6U	308	1JT1	262	1ALD	363
1GWU	306	1GQ8	318	1C8Z	265	1C0P	363
1PA2	306	1IZD	323	1AKO	268	1C3P	372
1BGP	309	1BXO	323	1R4X	273	1GW1	375
2IE7	318	1DS1	323	1F00	282	1ELJ	379
1AXN	323	1OEW	328	1NNH	293	1EDG	380
1LLP	343	1ODM	329	1HQ0	295	1BS0	383
1QAZ	351	1MZY	331	1H72	296	1BI5	388
1M15	356	1K5C	333	1BQB	301	1JHD	396
1YYD	357	1IA5	339	1IOW	306	1AYE	401
1KWF	363	1HG8	349	1GSA	314	1DLJ	402
1CA1	370	1H6L	353	1CS8	315	1HT6	404
1IOM	374	1IDK	359	1QLM	316	1KMJ	404
1YT3	375	1JTA	361	1NIJ	317	1EU8	407
2FUZ	377	1FNF	368	1M2R	327	1HDI	413
1V5C	386	1CA1	370	1B8P	327	16PK	415
1N40	394	1BHE	376	1B63	333	1ITX	419
1QHD	397	1CS6	382	1NE9	335	1D5T	433
1Z8O	402	1NOF	383	1M15	356	1F20	435
1DLJ	402	1F8E	388	1OWQ	361	1KFW	435
1H12	404	1JHD	396	1M7V	361	1E4I	447
2J07	419	1HT6	404	1C0P	363	1FF9	447
1KS8	433	1GPI	430	1NU5	369	1G94	448
1CSH	435	1MXG	435	1PBE	391	1BGV	449
1R89	437	1F20	435	1AYE	401	1J09	468
1J09	468	1G94	448	1PCA	402	1LFW	468
1GAI	472	1A8D	452	1EYN	418	1JAE	470
2FBA	492	1O6T	462	1ITX	419	1E43	483
2D5B	500	1JAE	470	1G6S	427	1LAM	484
1GKM	509	1AKL	470	1D5T	433	1HX0	495
1DL2	510	1OFL	480	1KFW	435	1E4M	499
1PFV	547	1E43	483	1NHS	446	1ESW	500
1Y0P	568	1HX0	495	1FF9	447	1DPE	507
1N5U	583	1GYC	499	1EWF	456	1EA5	532
1HS6	610	1HFU	500	1LFW	468	1EDQ	540
1QSA	618	1NKG	508	1AKL	470	1IAT	556
1G9G	629	1EI5	518	1N4W	498	1FEH	574
1KV9	664	1EDQ	540	1HP1	516	1CF3	581
1CB8	674	1HS6	610	1EDQ	540	1G5A	628
1N7O	721	1NR0	610	1FEH	574	1KWG	644
1X1I	752	1G5A	628	1CF3	581	1CXL	686
1RWH	754	1KWG	644	1HS6	610	1BF2	750
1YGE	839	1K3I	650	1QSA	618	1H16	759
1VLB	907	1KV9	664	1R1H	696	1KBL	872