



MÉTODOS DE CONTROLE ÓTIMO DOS PROCESSOS DINÂMICOS DO ENOVELAMENTO DE PROTEÍNAS

Simão Coutinho de Albuquerque Neto

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Rubem Pinto Mondaini

Rio de Janeiro
Julho de 2015

MÉTODOS DE CONTROLE ÓTIMO DOS PROCESSOS DINÂMICOS DO
ENVELAMENTO DE PROTEÍNAS

Simão Coutinho de Albuquerque Neto

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE
SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Rubem Pinto Mondaini, D.Sc.

Prof. Marco Antonio von Krüger, Ph.D.

Prof. Heraldo Luis Silveira de Almeida, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JULHO DE 2015

Albuquerque Neto, Simão Coutinho de

Métodos de Controle Ótimo dos Processos Dinâmicos do Enovelamento de Proteínas/Simão Coutinho de Albuquerque Neto. – Rio de Janeiro: UFRJ/COPPE, 2015.

XII, 38 p.: il.; 29,7cm.

Orientador: Rubem Pinto Mondaini

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2015.

Referências Bibliográficas: p. 37 – 38.

1. Controle ótimo. 2. Enovelamento de proteínas.
3. Dinâmica de macromoléculas. I. Mondaini, Rubem Pinto. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha família.

Agradecimentos

Aos meus pais, por todo apoio e carinho dados em todos os momentos de minha vida.

Agradeço ao meu orientador, Professor Rubem Mondaini, pelo curso de Controle Ótimo lecionado por ele na Escola Politécnica da UFRJ, pela oportunidade dada, pelo acompanhamento experiente das etapas necessárias à consecução desta dissertação e por todos os conselhos e incentivos que visam a mostrar-me o que significa o trabalho de um cientista.

Ele sempre me disse que o estágio de mestrado era definidor da escola de formação científica e um filtro de aptidões. Estas orientações continuam a manter o objetivo de obter resultados válidos perante a comunidade científica internacional quando da realização de uma futura tese de doutorado.

Aos meus amigos, que, apesar das reclamações sobre minha ausência nos últimos tempos, me deram sempre motivação. Em especial, agradeço a Felipe Ribeiro pelo auxílio com o \LaTeX .

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MÉTODOS DE CONTROLE ÓTIMO DOS PROCESSOS DINÂMICOS DO ENOVELAMENTO DE PROTEÍNAS

Simão Coutinho de Albuquerque Neto

Julho/2015

Orientador: Rubem Pinto Mondaini

Programa: Engenharia de Sistemas e Computação

O presente trabalho se utiliza de técnicas de controle ótimo, uma ferramenta matemática muito potente, baseadas no Princípio de Máximo de Pontryagin, para analisar dinâmicas da estrutura de proteínas. A partir de um modelo dinâmico para os átomos encontrados na espinha dorsal do polipeptídeo, analisam-se as consequências de consideração de uma dinâmica controlada por métodos de controle ótimo. São analisados dois casos. No primeiro, o termo inercial é considerado como pequeno diante do termo de fricção. No segundo caso, as diferenças entre as massas dos sítios atômicos são consideradas como sendo pequenas.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

OPTIMAL CONTROL METHODS OF DYNAMICAL PROCESSES OF
PROTEIN FOLDING

Simão Coutinho de Albuquerque Neto

July/2015

Advisor: Rubem Pinto Mondaini

Department: Systems Engineering and Computer Science

This work uses optimal control techniques, a very powerful mathematical tool based on Pontryagin's Maximum Principle, to analyze the dynamics of proteins' molecular structure. From a dynamical model for the polypeptide backbone's atoms, the consequences of the consideration of a controlled dynamic by optimal control methods are analyzed. Two different cases are analyzed. In the first, the inertial term is considered negligible compared to the friction term. In the second case, the differences between the atoms' sites are considered very small.

Sumário

Lista de Figuras	ix
Lista de Tabelas	xi
Lista de Abreviaturas	xii
1 Introdução	1
1.1 Organização	2
2 Enovelamento de Proteínas	3
2.1 Proteínas	3
2.2 Estrutura química	4
2.3 Aspectos conformacionais	7
3 Métodos usuais de controle aplicados à dinâmica de proteínas	10
4 Metodologia	12
4.1 Modelo dinâmico	12
4.2 Configuração do problema	17
4.2.1 Caso de termos inerciais desprezíveis diante do termo de fricção	19
4.2.2 Caso de pequena diferença entre as massas dos sítios atômicos	20
4.3 Equações Adimensionais	21
4.3.1 Caso de termos inerciais desprezíveis diante do termo de fricção	22
4.3.2 Caso de pequena diferença entre as massas dos sítios atômicos	23
5 Resultados	26
5.1 Caso de termos inerciais desprezíveis diante do termo de fricção . . .	26
5.2 Caso de pequena diferença entre as massas dos sítios atômicos	28
5.3 Discussão sobre os resultados	35
6 Conclusão	36
Referências Bibliográficas	37

Lista de Figuras

2.1	Processo de síntese de proteínas	3
2.2	Estrutura de um aminoácido.	4
2.3	Ligação peptídica entre dois aminoácidos.	6
2.4	Polipeptídeo com N aminoácidos.	6
2.5	A hélice- α	8
2.6	A folha- β	8
3.1	Modelo de contas esféricas centradas em C^α	11
4.1	Exemplo de um peptídeo com os vetores unitários.	15
4.2	Exemplo de combinações com quatro caixas distintas, uma bola azul e uma bola vermelha.	21
5.1	Comparação dos estados com termos inerciais desprezíveis diante do termo de fricção.	27
5.2	Comparação dos controles com termos inerciais desprezíveis diante do termo de fricção.	27
5.3	t_F máximo para o caso $p = 1, q = 2$	28
5.4	t_F máximo para o caso $p = 2, q = 2$	28
5.5	Comparação dos estados com $\Delta m \approx 0$ e $a = 1$	29
5.6	Comparação dos controles com $\Delta m \approx 0$ e $a = 1$	29
5.7	t_F para o caso $p = 1, q = 2$	29
5.8	t_F para o caso $p = 2, q = 2$	30
5.9	Gráfico do estado l_1 para $a = 0.597$	30
5.10	Gráfico do controle u_1 para $a = 0.597$	30
5.11	t_{1F} para $a = 0.597$	31
5.12	Valor de t em que l_1 cruza o eixo para $a = 0.597$	31
5.13	Gráfico do estado l_2 para $a = 0.965$	32
5.14	Gráfico do controle u_2 para $a = 0.965$	32
5.15	t_{2F} para $a = 0.965$	32
5.16	Gráfico do estado l_1 para $a = 1.937$	33
5.17	Gráfico do controle u_1 para $a = 1.937$	33

5.18	t_{1F} para $a = 1.937$.	33
5.19	Gráfico do estado l_2 para $a = 2.094$.	34
5.20	Gráfico do controle u_2 para $a = 2.094$.	34
5.21	t_{2F} para $a = 2.094$.	34

Lista de Tabelas

1.1	Constituição das células excetuando a água	1
2.1	Os 20 tipos comuns de aminoácidos.	5
5.1	O comportamento detectado nos dois casos, $p = 1, q = 2$ e $p = 2, q = 2$	35

Lista de Abreviaturas

C^α	- Carbono- α
MD	- Molecular Dynamics
CG	- Coarse-grained

Capítulo 1

Introdução

Proteínas são macromoléculas de extrema importância pois exercem diversos tipos de funções nos organismos de todos os seres vivos. Após a água, são as moléculas em maior quantidade na composição das células (Tabela 1.1). A evolução da vida está diretamente relacionada à evolução na complexidade das proteínas e de sua capacidade de exercer novas funcionalidades.

Tabela 1.1: Constituição das células excetuando a água. Adaptado de [1]

Composto	Porcentagem
Proteínas	50
Carboidratos	15
Ácidos nucleicos	15
Lipídios	10
Outros	10

O chamado *problema de enovelamento de proteínas* está relacionado com a predição do estado enovelado da proteína a partir de sua sequência de aminoácidos, e da determinação da função protéica através da observação da conformação tridimensional. O estudo do processo físico pode auxiliar na descoberta dos fatores responsáveis por erros ocasionados na conformação de determinadas proteínas que acarretam em doenças e, muito além, no desenvolvimento das Ciências como um todo.

Apesar da criação de proteínas a partir da manipulação de sequências de aminoácidos já não ser uma grande dificuldade, o conhecimento prévio da conformação resultante permanece como um desafio. Uma vez que esta dificuldade tenha sido suprimida, a criação de proteínas sintéticas com propriedades específicas, não encontradas no meio em que vivemos, tornar-se-á uma tarefa consideravelmente mais simples.

Devido à grande importância das proteínas, diversos estudos, utilizando diferen-

tes tipos de técnicas, já foram realizados com o objetivo de destrinchar o problema do enovelamento. A literatura cobrindo todos estes aspectos é extremamente vasta, então selecionamos algumas referências que tratassem o problema com métodos de controle.

Para um modelo de representação dos elementos presentes na espinha dorsal de uma proteína, a dinâmica é analisada utilizando métodos de controle ótimo. São adotados e comparados o caso de massas associadas aos sítios atômicos desprezíveis, muito presente na literatura, e o caso de pequena diferença entre as massas dos sítios atômicos [2].

1.1 Organização

No Capítulo 2 é dada uma explicação resumida sobre o que é uma proteína, abordando sua estrutura química e seu aspecto conformacional.

Alguns métodos usais para a análise de problemas de enovelamento de proteínas são apresentados no Capítulo 3.

No Capítulo 4 é mostrado a configuração do problema de controle ótimo. Para isto, mostramos o modelo dinâmico adotado para a análise e fazemos um processo de adimensionalização do sistema de equações.

O Capítulo 5 apresenta representações gráficas de alguns estados elementares assim como seus controles.

No Capítulo 6 é apresentada a conclusão do estudo e possíveis trabalhos futuros são discutidos.

Capítulo 2

Enovelamento de Proteínas

2.1 Proteínas

Todos os seres vivos são formados pelos mais variados tipos de células, mas estas se utilizam dos mesmos mecanismos para realizar as mais básicas funções. No DNA se encontram as informações para a execução da grande maioria dos processos químicos da célula. Porém, para isto, ele deve iniciar a produção de outros polímeros: RNAs e proteínas. Segmentos da sequência de DNA são transcritos em moléculas menores (RNA) que, posteriormente, no processo de tradução, dão origem às proteínas (Figura 2.1). Tanto DNA, RNA e proteínas são compostos pelos mesmos seis elementos: hidrogênio, carbono, nitrogênio, oxigênio, enxofre e fósforo.

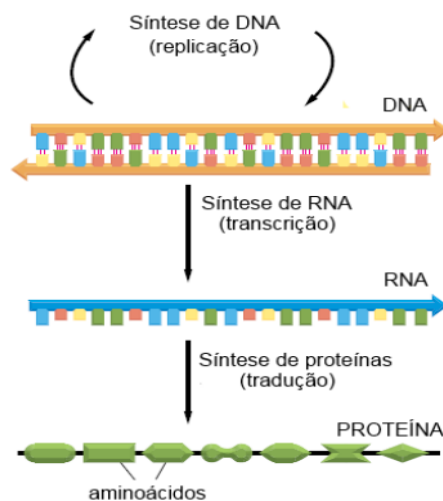


Figura 2.1: Processo de síntese de proteínas. Adaptado de [3]

As proteínas são as moléculas funcionais principais na manutenção da vida e executam quase todas as funções da célula. Apresentam diversas funções específicas que variam de acordo com suas sequências de aminoácidos, dentre elas a síntese de novas moléculas de DNA, cuja informação genética é utilizada na origem do RNA

e de proteínas. A ocorrência de mutações nas proteínas pode fazer com que elas percam suas funções e acarrete no surgimento de doenças. O estudo de proteínas tem como objetivo prever suas estruturas e atividades, de forma a permitir o desenvolvimento de drogas que minimizem problemas de saúde.

2.2 Estrutura química

As moléculas de proteínas, assim como as de DNA e RNA, são cadeias de polímeros não ramificadas, mas ao contrário destes, seus monômeros, os aminoácidos, são de 20 tipos diferentes ao invés de quatro – adenina (A), citosina (C), guanina (G) e timina (T) no DNA e A, C, G e uracila (U) no RNA. Os aminoácidos são compostos variados, cada qual com sua propriedade química, porém são construídos a partir da mesma estrutura (Figura 2.2): nas extremidades um grupo amina (NH_2) e um grupo carboxila (COOH) conectados a um átomo de carbono central (chamado carbono alfa, C^α), que ainda tem como ligações um átomo de hidrogênio e uma cadeia lateral, representada pela letra R, que diferencia e identifica cada aminoácido específico (Tabela 2.1), responsável pela variedade química entre eles.

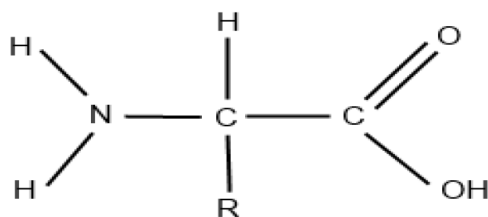












Figura 2.2: Estrutura de um aminoácido.

Cada uma das muitos milhares de moléculas existentes de proteínas é formada por uma sequência longa e única de aminoácidos que, por sua vez, é responsável pela função específica de cada proteína. Têm geralmente entre 50 e 2000 aminoácidos de comprimento [3], mas podem variar entre 25 a mais de 10000 [5]. Os aminoácidos se ligam uns aos outros de uma maneira padrão, chamada ligação peptídica (covalente), o que implica numa direcionalidade definitiva na molécula: um grupo amina numa ponta e um grupo carboxila na outra (Figura 2.3). A sequência de aminoácidos que forma o polímero é chamada de estrutura primária da proteína (Figura 2.4), enquanto que à sequência NC^αC repetida ao longo da cadeia polipeptídica é dado o nome de espinha dorsal.

Para exercer sua função biológica, a molécula de proteína passa por um processo onde se enrola e se dobra sobre si mesma assumindo um formato tridimensional específico. A configuração resultante, chamada estado enovelado da proteína, é diretamente relacionada à sequência de aminoácidos do polímero. Muitas se enovelam

Tabela 2.1: Os 20 tipos comuns de aminoácidos. Adaptado de [1], [3] e [4]

Aminoácido	Códigos		Cadeia Lateral	Polaridade
	3 letras	1 letra		
Alanina	Ala	A	$-\text{CH}_3$	apolar
Cisteína	Cys	C	$-\text{CH}_2\text{SH}$	apolar
Ácido aspártico	Asp	D	$-\text{CH}_2\text{O}_2^-$	negativa
Ácido glutâmico	Glu	E	$-(\text{CH}_2)_2\text{O}_2^-$	negativa
Fenilalanina	Phe	F	$-\text{CH}_2-$ 	apolar
Glicina	Gly	G	$-\text{H}$	apolar
Histidina	His	H	$-\text{CH}_2-$ 	positiva
Isoleucina	Ile	I	$-\text{CH}$ 	apolar
Lisina	Lys	K	$-(\text{CH}_2)_4\text{NH}_3^+$	positiva
Leucina	Leu	L	$-\text{CH}_2-\text{CH}$ 	apolar
Metionina	Met	M	$-(\text{CH}_2)_2\text{SCH}_3^+$	apolar
Asparagina	Asn	N	$-\text{CH}_2\text{C}(=\text{O})\text{NH}_2$	neutra
Prolina	Pro	P		apolar
Glutamina	Gln	Q	$-(\text{CH}_2)_2\text{C}(=\text{O})\text{NH}_2$	neutra
Arginina	Arg	R	$-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}-\text{C}$ 	positiva
Serina	Ser	S	$-\text{CH}_2\text{OH}$	neutra
Treonina	Thr	T	$-\text{CH}$ 	neutra
Valina	Val	V	$-\text{CH}$ 	apolar
Triptofano	Trp	W	$-\text{CH}_2-$ 	apolar
Tirosina	Tyr	Y	$-\text{CH}_2-$ 	neutra

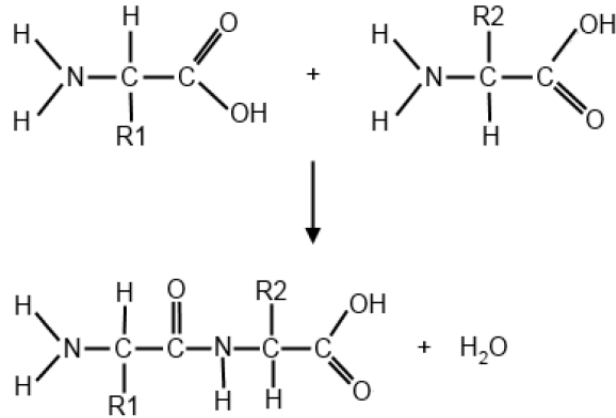


Figura 2.3: Ligação peptídica entre dois aminoácidos.

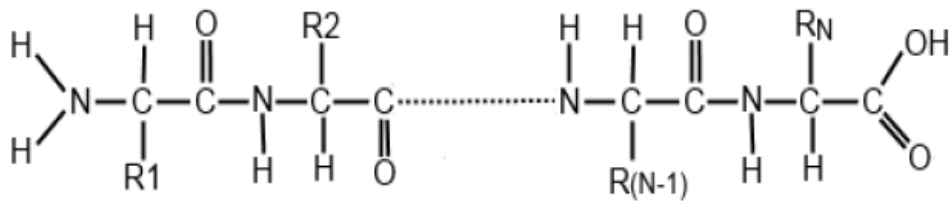


Figura 2.4: Polipeptídeo com N aminoácidos.

sem ajuda enquanto outras têm seu progresso acompanhadas por proteínas auxiliares chamadas chaperonas, que dão uma maior confiabilidade ao processo e servem como um controle de qualidade.

Há alguns fatores que limitam as formas possíveis que uma proteína pode assumir. Além da impossibilidade de sobreposição dos átomos, a polaridade dos aminoácidos, ligações de hidrogênio, atrações eletrostáticas e interações de van der Waals são determinantes ao arranjo tridimensional final. Estas três últimas são consideradas fracas se comparadas às ligações covalentes responsáveis por formar a cadeia polipeptídica, porém um grande número combinado de tais ligações resultam em força suficiente para manter a estabilidade em cada estado de enovelamento. A polaridade é também um fator de extrema importância, uma vez que os aminoácidos hidrofílicos (polares) tendem a ficar na parte externa da molécula, evitando que os hidrofóbicos (apolares), que costumam se agrupar no interior, tenham contato com a água do meio circundante. O estudo das prováveis disposições dos resíduos polares auxiliam no reconhecimento dos estados nativos enovelados [6].

Proteínas apresentam normalmente apenas uma conformação estável que, geralmente, é a que minimiza sua energia livre (o mínimo global pode estar localizado em uma região “inacessível” [7]). Contudo, ao interagir com outras moléculas, seu formato pode sofrer pequenas alterações para exercer devidamente sua função. Estas transições conformacionais se dão através da transição do sistema de um estado

metaestável do potencial para outro [5, 8, 9]. Todas as proteínas se ligam a outras moléculas e a interação física com elas é que determina sua propriedade biológica. Cada uma se vincula a substâncias específicas que são denominadas ligantes para essa proteína.

Estudos realizados por biólogos com a utilização de solventes que rompem as interações não covalentes que mantêm a proteína enovelada unida, indicaram que a sequência de aminoácidos contém toda a informação necessária para chegar ao seu estado enovelado [3]. Com o tratamento, a proteína é convertida em uma cadeia flexível e, após a retirada do solvente, frequentemente retorna a sua conformação de forma espontânea.

Diversas reações bioquímicas e processo biológicos dependem da flexibilidade e de mudanças estruturais nas proteínas. Acredita-se que tais mudanças são regidas por modos de vibração de baixa frequência que se estendem através da proteína. Trabalhos recentes [10, 11] encontraram que os modos vibracionais subamortecidos envolvem os movimentos estruturais da proteína. Estes modos subamortecidos teriam também um papel relevante no percurso pela superfície de energia potencial em busca da conformação mais estável, ou seja, as mudanças conformacionais estariam associadas a alterações nos movimentos vibracionais intramoleculares das proteínas.

2.3 Aspectos conformacionais

Apesar da conformação global de cada proteína ser única, alguns padrões regulares e repetitivos são observados em grande parte delas. Removendo as cadeias laterais, ficam visíveis elementos chamados hélices- α e folhas- β que resultam de ligações de hidrogênio entre grupos $N - H$ e $C = O$ na espinha dorsal do polipeptídeo. Esta organização é chamada de estrutura secundária da proteína, enquanto a conformação completa (com as cadeias laterais) é designada como estrutura terciária. Uma molécula mais complexa, formada por idênticas ou diferentes cadeias polipeptídicas que se encaixam, tem sua estrutura completa referida como estrutura quaternária.

A hélice- α é uma estrutura estável bastante comum (corresponde a cerca de um terço dos resíduos de aminoácidos) que se forma quando uma cadeia polipeptídica se torce em torno de si mesma (Figura 2.5). Cada grupo $C = O$ faz uma ligação de hidrogênio com o grupo $N - H$ do quarto aminoácido de distância. Cada volta completa da hélice contém 3,6 aminoácidos e, cada hélice consiste em média de quatro voltas.

Folhas- β são estruturas bastante rígidas, mantidas por ligações de hidrogênio entre cadeias polipeptídicas vizinhas (Figura 2.6). Existem duas formações possíveis: paralelas, onde as cadeias vizinhas alinham-se na mesma direção, e antiparalelas,

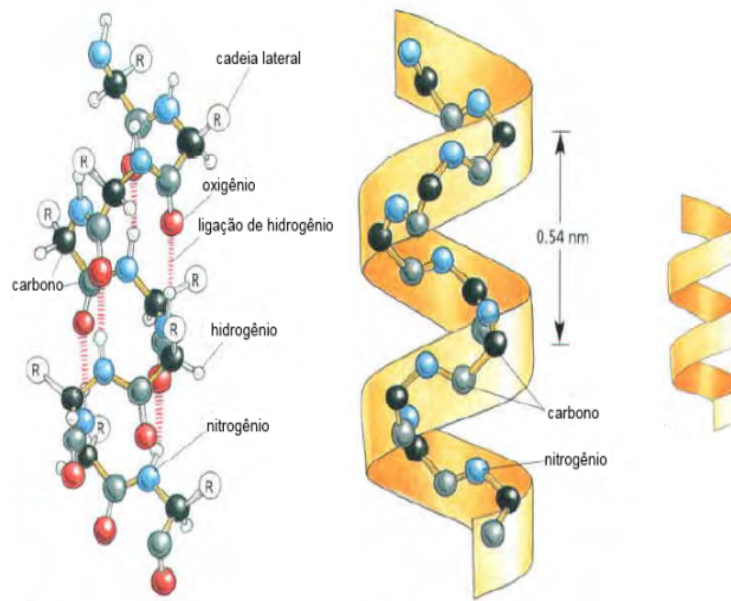


Figura 2.5: A hélice- α . Da esquerda para a direita: Estrutura completa com todos os átomos, estrutura contendo apenas os átomos da espinha dorsal, representação simbólica comumente utilizada. Adaptado de [3]

com cada seção da cadeia alternando a sua direção em relação aos seus vizinhos.

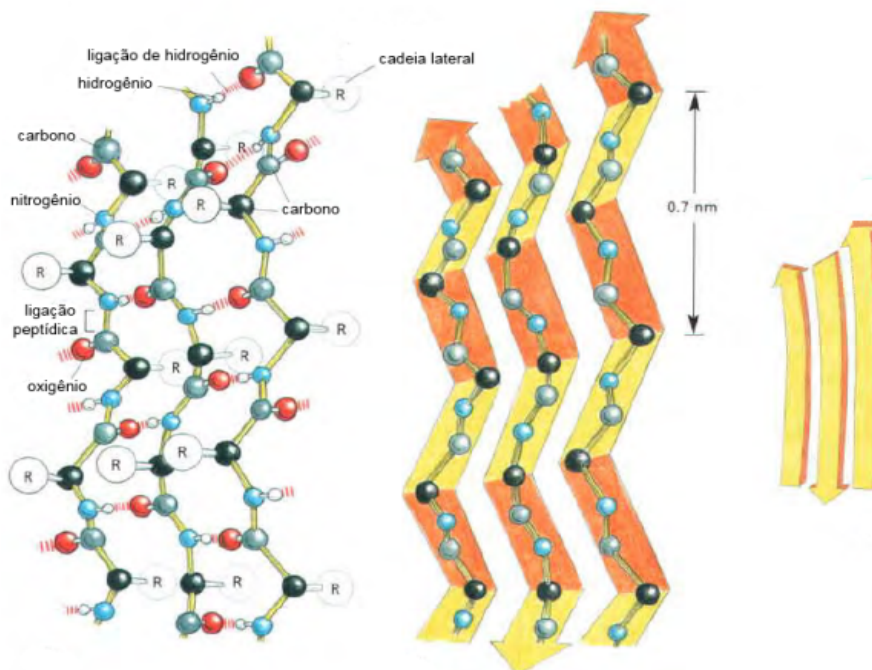


Figura 2.6: A folha- β . Da esquerda para a direita: Estrutura completa com todos os átomos, estrutura contendo apenas os átomos da espinha dorsal, representação simbólica comumente utilizada. Adaptado de [3]

O processo de enovelamento da cadeia polipeptídica não se dá de uma extremidade à outra, mas sim por grupos separados, denominados *domínios proteicos*, que se enovelam em estruturas estáveis de forma independente. Para a determinação

dos domínios estruturais, o estado enovelado da proteína deve ser conhecido. Uma proteína pode conter de um à várias dezenas de domínios. Estes, por sua vez, têm geralmente entre 40 e 350 aminoácidos [3].

O número de domínios está relacionado à complexidade da função realizada pela proteína. Apresentam mais relações intra-atômicas do que com outros grupos, e são as unidades fundamentais na estrutura da proteína pois contêm as suas características básicas e podem, em alguns casos, funcionar sozinhos [3, 12]. Acredita-se que a evolução na complexidade das proteínas seja consequência de junções e/ou troca de domínios em processo chamado *embaralhamento de domínios*.

Proteínas com estruturas tridimensionais semelhantes são agrupadas em famílias. Suas sequências de aminoácidos não são necessariamente próximas, chegando a divergir bastante, e suas funções também são frequentemente distintas. Estes fatores seriam devidos às trocas e mutações ocasionados no decorrer da evolução, e todos os integrantes da família seriam descendentes de um ancestral em comum.

Capítulo 3

Métodos usuais de controle aplicados à dinâmica de proteínas

Há na literatura a utilização de diversos métodos para o estudo da dinâmica do enovelamento de proteínas. Desde os mais precisos, que precisam de supercomputadores e cujas simulações são de longa duração, até os mais simples, que muitas vezes se mostram suficientes para compreender alguns aspectos do processo de enovelamento.

Métodos mais precisos como o de Dinâmicas Moleculares (*Molecular Dynamics* - MD) podem incluir todas as interações atômicas da proteína na modelagem, até mesmo as entre o solvente, o que os tornam muito mais complexos conforme o número de resíduos aumenta. Porém, tamanha precisão demanda um esforço computacional muito além do que a tecnologia atual é capaz de fornecer, limitando as simulações em tempo a real a frações muito curtas do processo.

Modelos simplificados ou grosseiros (ou de granulação grosseira numa tradução mais literal de *coarse-grained* - CG) são bastante comuns pois reduzem consideravelmente os tempos de simulação e, muitas vezes, são capazes de fornecer informações estruturais relevantes do processo de enovelamento. Uma das formas mais comuns é a modelagem de proteínas como cadeias lineares de contas esféricas [13, 14]. Estas contas podem tanto representar os átomos da espinha dorsal do polipeptídeo quanto os aminoácidos (com o centro da esfera definido pela posição do C^α). A utilização de modelos que se utilizam apenas das coordenadas do C^α são muito frequentes na literatura e aparecem até na identificação de domínios estruturais [15].

Como exemplo de modelos simplificados, os chamados modelos *Go-type*, são bastante utilizados para o estudo das trajetórias de enovelamento. Conhecendo previamente o estado enovelado da proteína (disponível em bancos de dados de livre acesso), o modelo prioriza as interações nativas, de forma que as contas esféricas sabem em qual posição devem estar no fim da simulação, mas não sabem o caminho que irão percorrer. A superfície de energia apresenta formato afunilado ao invés de rugosa como as obtidas por simulações mais “realísticas” [16], mas isto seria uma

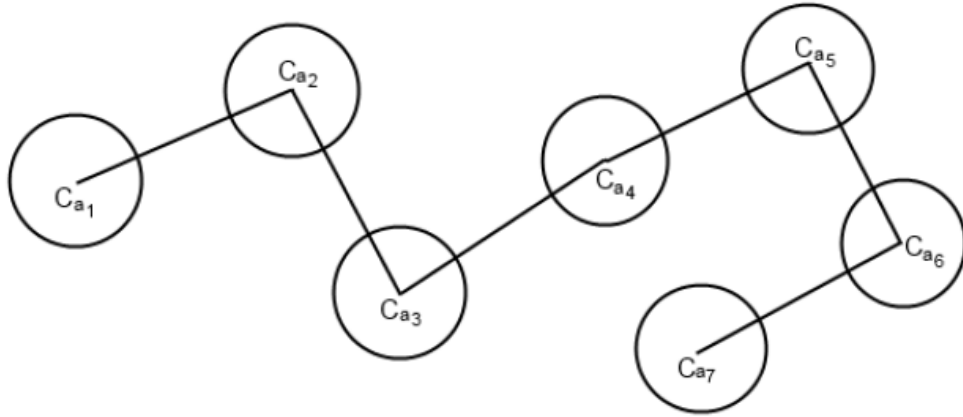


Figura 3.1: Modelo de contas esféricas centradas em C^α

característica positiva pois em proteínas reais algumas interações parecem ser favorecidas, levando a caminhos preferenciais de enovelamento (existe uma discordância sobre se o fator crucial seriam as interações hidrofóbicas [17] ou hidrofílicas [7]). Conforme Levinthal apontou, se as proteínas vasculhassem todas as configurações possíveis na busca do seu estado nativo, o processo levaria mais do que a idade do universo. Como as sequências de aminoácidos “escolhem” os caminhos mais favoráveis para chegar aos seus estados enovelados nativos de forma tão rápida é uma das grandes questões que permeiam a biologia molecular.

Modelos com “grades” (*lattice models*) são ainda mais simples, pois além dos aminoácidos serem considerados como contas esféricas centradas em C^α , suas posições são restritas à grades cúbicas. Um dos mais populares, o modelo **HP**, utiliza apenas dois tipos de contas esféricas, escolhidas conforme a polaridade dos aminoácidos: **H** (hidrofóbicas/apolares) e **P** (hidrofílicas/polares). Simulações Monte Carlo são frequentemente utilizadas em modelos com “grades” [18].

Em [19], um modelo combinando modelagem CG e MD é adotado para estudar as trajetórias do enovelamento de uma proteína específica, cuja conformação final é conhecida. Usando um modelo de contas centradas em C^α , os contatos são atualizados a cada passo do processo, levando ou deixando de levar em consideração interações não nativas conforme elas ocorram ou deixem de ocorrer. As posições iniciais e finais das contas, e as forças finais são conhecidas. No modelo dinâmico, os termos contendo as massas são supostos negligenciáveis e o controle responsável por levar a proteína de um estrutura inicial não enovelada para o seu estado nativo é quadrático. O modelo CG faz de forma rápida o cálculo das trajetórias das contas durante um pequeno intervalo de tempo do tempo total de simulação. O resultado é enviado ao modelo MD que o refina e o devolve ao CG, atualizando a curva de potencial deste. O ciclo é repetido diversas vezes até alcançar o estado nativo.

Capítulo 4

Metodologia

4.1 Modelo dinâmico

Como objeto de estudo, supomos um caso em que inicialmente uma proteína se encontre em um estado enovelado correspondente a um mínimo de energia e , devido a algum fator externo, como por exemplo a interação com outra molécula, ocorram pequenas vibrações que levem o sistema a outra posição de equilíbrio. Esta transição resultaria em uma pequena alteração na conformação do sistema molecular.

O modelo adotado neste trabalho se utiliza de uma forma simplificada de representação da molécula de proteína. Apenas os átomos de carbono e nitrogênio presentes na espinha dorsal da cadeia polipeptídica são levados em consideração.

Caracterizamos o enovelamento como um processo dinâmico e, para a obtenção das equações de movimento do sistema, utilizamos do formalismo lagrangiano. Essa escolha se justifica pela simplificação no tratamento do sistema.

Considerando que os estados metaestáveis correspondam a mínimos locais adjacentes no poço de potencial, porém com energias levemente diferentes, implica que a variação de energia necessária para a transição do estado metaestável de maior energia para o de menor energia seja menor que a variação da transição do estado de menor energia para o de maior energia. Por exemplo, supomos a existência de dois mínimos com energias E_1 e E_2 , onde:

$$E_1 > E_2 ,$$

e com um máximo E_M entre eles. A variação de energia necessária para o sistema se deslocar do primeiro para o segundo estado é da seguinte forma:

$$\Delta E_{1,2} = E_M - E_1 ,$$

enquanto que no sentido contrário temos:

$$\Delta E_{2,1} = E_M - E_2.$$

Logo,

$$\Delta E_{2,1} > \Delta E_{1,2}.$$

Como podemos ver através deste exemplo, o trabalho realizado para levar o sistema de um mínimo para o outro depende da situação: se a transição ocorre do maior estado metaestável para o menor ou o contrário. Portanto, concluímos que o sistema é dissipativo.

As equações de movimento devem satisfazer a seguinte equação de Lagrange:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_j} \right) - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_j} = \mathbf{F}_j \quad \forall j, \quad (4.1)$$

onde \mathcal{L} é a função de Lagrange, \mathbf{r}_j o vetor posição do j-ésimo sítio atômico em relação a um ponto qualquer fixo no espaço, \mathbf{F}_j são forças generalizadas e $(\dot{})$ denota derivação em relação ao tempo. A função de Lagrange é definida como:

$$\mathcal{L} = T - V,$$

onde T é a energia cinética total do sistema e V a energia potencial advinda de forças conservativas. As interações entre os pares de átomos ligados são aqui consideradas como forças elásticas supostamente harmônicas. Desta forma, a energia cinética pode ser escrita como:

$$T = \frac{1}{2} \sum_j m_j \|\dot{\mathbf{r}}_j\|^2,$$

e a energia potencial:

$$V = \frac{1}{2} \kappa \sum_j \|\mathbf{r}_j - \mathbf{R}(t_0)\|^2.$$

Resultando em

$$\mathcal{L} = \frac{1}{2} \sum_j m_j \|\dot{\mathbf{r}}_j\|^2 - \frac{1}{2} \kappa \sum_j \|\mathbf{r}_j - \mathbf{R}(t_0)\|^2 \quad (4.2)$$

onde m_j é a massa do j-ésimo átomo, κ a constante elástica e t_0 o tempo inicial em que a proteína se encontra enovelada. Aqui, $\|\cdot\|$ denota a norma de um vetor, cujas componentes são as suas coordenadas cartesianas. Assim, a norma de um vetor \mathbf{r} qualquer é definida como:

$$\|\mathbf{r}\| = \sqrt{r_x^2 + r_y^2 + r_z^2}$$

\mathbf{R} é o vetor posição do centro de massa da proteína:

$$\mathbf{R}(t_0) = \frac{\sum_k m_k \mathbf{r}_k(t_0)}{\sum_k m_k}$$

Como é suposto o caso em que a proteína se encontra enovelada em um mínimo de energia e transita para outro estado através de pequenas oscilações, o centro de massa é considerado como estando em repouso.

As forças \mathbf{F}_j são decompostas da seguinte forma:

$$\mathbf{F}_j = -\frac{\partial \mathcal{F}}{\partial \dot{\mathbf{r}}_j} + \mathbf{U}_j, \quad (4.3)$$

onde \mathcal{F} é a função dissipação de Rayleigh:

$$\mathcal{F} = \frac{1}{2} \gamma \sum_i \|\dot{\mathbf{r}}_i\|^2, \quad (4.4)$$

com γ representando a constante de atrito do meio, e \mathbf{U} é o vetor contendo o restante das forças. Substituindo a equação (4.3) na equação de Lagrange (4.1), temos:

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_j} \right) - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_j} + \frac{\partial \mathcal{F}}{\partial \dot{\mathbf{r}}_j} = \mathbf{U}_j. \quad (4.5)$$

Substituindo (4.4) e (4.2) em (4.5) e realizando as derivações, obtemos:

$$m_j \ddot{\mathbf{r}}_j + \kappa (\mathbf{r}_j - \mathbf{R}(t_0)) + \gamma \dot{\mathbf{r}}_j = \mathbf{U}_j. \quad (4.6)$$

Colocando em termos dos átomos de carbono e nitrogênio da espinha dorsal da proteína, temos:

$$\begin{cases} m_{N_j} \ddot{\mathbf{r}}_{N_j} = -\gamma \dot{\mathbf{r}}_{N_j} - \kappa (\mathbf{r}_{N_j} - \mathbf{R}(t_0)) + \mathbf{U}_{N_j} \\ m_{C_j^\alpha} \ddot{\mathbf{r}}_{C_j^\alpha} = -\gamma \dot{\mathbf{r}}_{C_j^\alpha} - \kappa (\mathbf{r}_{C_j^\alpha} - \mathbf{R}(t_0)) + \mathbf{U}_{C_j^\alpha} \\ m_{C_j} \ddot{\mathbf{r}}_{C_j} = -\gamma \dot{\mathbf{r}}_{C_j} - \kappa (\mathbf{r}_{C_j} - \mathbf{R}(t_0)) + \mathbf{U}_{C_j} \end{cases}$$

Como as massas dos nitrogênios localizados no j -ésimo e no do $(j+1)$ -ésimo sítios atômicos são as mesmas, então as massas independem do índice j . O mesmo é válido para os átomos de carbono, não havendo também diferença em relação às massas

dos C^α 's, ou seja, $m_{C^\alpha} = m_C$. Logo:

$$\begin{cases} m_N \ddot{\mathbf{r}}_{N_j} = -\gamma \dot{\mathbf{r}}_{N_j} - \kappa (\mathbf{r}_{N_j} - \mathbf{R}(t_0)) + \mathbf{U}_{N_j} \\ m_C \ddot{\mathbf{r}}_{C_j^\alpha} = -\gamma \dot{\mathbf{r}}_{C_j^\alpha} - \kappa (\mathbf{r}_{C_j^\alpha} - \mathbf{R}(t_0)) + \mathbf{U}_{C_j^\alpha} \\ m_C \ddot{\mathbf{r}}_{C_j} = -\gamma \dot{\mathbf{r}}_{C_j} - \kappa (\mathbf{r}_{C_j} - \mathbf{R}(t_0)) + \mathbf{U}_{C_j} \end{cases} \quad (4.7)$$

Definimos os módulos l_j dos comprimentos de ligações entre os átomos da espinha dorsal e os vetores unitários $\hat{\mathbf{u}}_j$, $\hat{\mathbf{v}}_j$ e $\hat{\mathbf{w}}_j$ na direção das ligações, de forma que podemos escrever:

$$\begin{aligned} \mathbf{r}_{C_j} - \mathbf{r}_{C_j^\alpha} &= l_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j, \\ \mathbf{r}_{N_{j+1}} - \mathbf{r}_{C_j} &= l_{N_{j+1}, C_j} \hat{\mathbf{v}}_j, \\ \mathbf{r}_{C_{j+1}^\alpha} - \mathbf{r}_{N_{j+1}} &= l_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j. \end{aligned}$$

A figura 4.1 mostra a representação da estrutura de um peptídeo com os vetores unitários. Nesta representação, os átomos de nitrogênio do grupo amina e de carbono da carboxila das extremidades não são considerados, e temos $j = 1, \dots, n-1$, onde n é igual ao número de resíduos.

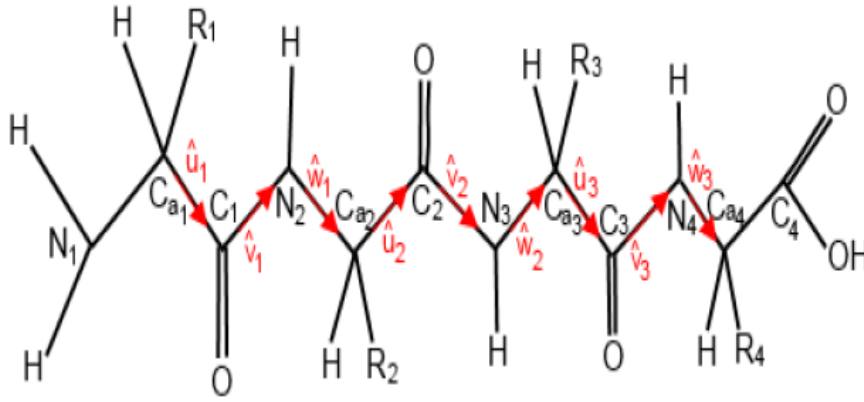


Figura 4.1: Exemplo de um peptídeo com os vetores unitários. Os grupos amina e carboxila das extremidades não participam deste modelo. Aqui, $\hat{\mathbf{u}}_j$, $\hat{\mathbf{v}}_j$ e $\hat{\mathbf{w}}_j$ são os vetores unitários de $\mathbf{r}_{C_j} - \mathbf{r}_{C_j^\alpha}$, $\mathbf{r}_{N_{j+1}} - \mathbf{r}_{C_j}$ e $\mathbf{r}_{C_{j+1}^\alpha} - \mathbf{r}_{N_{j+1}}$, respectivamente

O vetor de forças \mathbf{U} , assim como as funções elásticas, é suposto de tal forma que a sua ação se dê também apenas ao longo das ligações covalentes. Podemos então definir o módulo u , referente às forças:

$$\begin{aligned} \mathbf{U}_{C_j} - \mathbf{U}_{C_j^\alpha} &= u_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j, \\ \mathbf{U}_{N_{j+1}} - \mathbf{U}_{C_j} &= u_{N_{j+1}, C_j} \hat{\mathbf{v}}_j, \\ \mathbf{U}_{C_{j+1}^\alpha} - \mathbf{U}_{N_{j+1}} &= u_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j. \end{aligned}$$

O intervalo de tempo ΔT em que ocorre a transição de um estado metaestável para outro é considerado muito rápido se comparado com o tempo em que o sistema permanece em um mínimo. Supomos então que os vetores unitários $\hat{\mathbf{u}}_j$, $\hat{\mathbf{v}}_j$ e $\hat{\mathbf{w}}_j$ não variam durante este intervalo. Logo, as condições iniciais:

$$\begin{aligned}\dot{\hat{\mathbf{u}}}_j(t_0) &= 0, \\ \dot{\hat{\mathbf{v}}}_j(t_0) &= 0, \\ \dot{\hat{\mathbf{w}}}_j(t_0) &= 0\end{aligned}$$

são mantidas durante todo o intervalo ΔT em que ocorre o movimento, ou seja:

$$\begin{aligned}\dot{\hat{\mathbf{u}}}_j(t_0 \leq t \leq t_0 + \Delta T) &= 0, \\ \dot{\hat{\mathbf{v}}}_j(t_0 \leq t \leq t_0 + \Delta T) &= 0, \\ \dot{\hat{\mathbf{w}}}_j(t_0 \leq t \leq t_0 + \Delta T) &= 0\end{aligned}$$

garantindo que os movimentos se dão ao longo das arestas.

Dada uma proteína contendo n resíduos, o sistema (4.7) pode ser reescrito em função das diferenças entre os átomos conectados da espinha dorsal:

$$\begin{cases} m_C \ddot{l}_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j = -\gamma \dot{l}_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j - \kappa l_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j + u_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j \\ m_N \ddot{\mathbf{r}}_{N_{j+1}} - m_C \ddot{\mathbf{r}}_{C_j} = -\gamma \dot{l}_{N_{j+1}, C_j} \hat{\mathbf{v}}_j - \kappa l_{N_{j+1}, C_j} \hat{\mathbf{v}}_j + u_{N_{j+1}, C_j} \hat{\mathbf{v}}_j \\ m_C \ddot{\mathbf{r}}_{C_{j+1}^\alpha} - m_N \ddot{\mathbf{r}}_{N_{j+1}} = -\gamma \dot{l}_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j - \kappa l_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j + u_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j \end{cases} \quad (4.8)$$

Dois casos foram propostos a serem analisados:

1. Considerando o termo inercial como pequeno diante do termo de fricção. Como γ pequeno e grande inibem consideravelmente o enovelamento, o termo em γ deve ser suficientemente grande para que o termo de inércia possa ser desprezado em relação a ele [13, 19, 20]:

$$\begin{aligned} m_C \ddot{\mathbf{r}}_{C_j^\alpha} &\approx 0, \quad m_C \ddot{\mathbf{r}}_{C_j} \approx 0, \quad m_N \ddot{\mathbf{r}}_{N_j} \approx 0 \\ \begin{cases} \gamma \dot{l}_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j = (-\kappa l_{C_j, C_j^\alpha} + u_{C_j, C_j^\alpha}) \hat{\mathbf{u}}_j \\ \gamma \dot{l}_{N_{j+1}, C_j} \hat{\mathbf{v}}_j = (-\kappa l_{N_{j+1}, C_j} + u_{N_{j+1}, C_j}) \hat{\mathbf{v}}_j \\ \gamma \dot{l}_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j = (-\kappa l_{C_{j+1}^\alpha, N_{j+1}} + u_{C_{j+1}^\alpha, N_{j+1}}) \hat{\mathbf{w}}_j \end{cases} \end{aligned} \quad (4.9)$$

2. Diferença insignificante entre as massas dos sítios atômicos:

$$m_C - m_N = \Delta m \approx 0$$

$$m_N \approx m_C = m$$

$$\begin{cases} m\ddot{l}_{C_j, C_j^\alpha} \hat{\mathbf{u}}_j = (-\gamma \dot{l}_{C_j, C_j^\alpha} - \kappa l_{C_j, C_j^\alpha} + u_{C_j, C_j^\alpha}) \hat{\mathbf{u}}_j \\ m\ddot{l}_{N_{j+1}, C_j} \hat{\mathbf{v}}_j = (-\gamma \dot{l}_{N_{j+1}, C_j} - \kappa l_{N_{j+1}, C_j} + u_{N_{j+1}, C_j}) \hat{\mathbf{v}}_j \\ m\ddot{l}_{C_{j+1}^\alpha, N_{j+1}} \hat{\mathbf{w}}_j = (-\gamma \dot{l}_{C_{j+1}^\alpha, N_{j+1}} - \kappa l_{C_{j+1}^\alpha, N_{j+1}} + u_{C_{j+1}^\alpha, N_{j+1}}) \hat{\mathbf{w}}_j \end{cases} \quad (4.10)$$

Ambos os casos resultam em sistemas cujas equações se dão apenas ao longo das ligações covalentes. Os conjuntos de equações (4.9) e (4.10) são reescritos, reespectivamente, como:

$$\begin{cases} \gamma \dot{l}_{C_j, C_j^\alpha} = -\kappa l_{C_j, C_j^\alpha} + u_{C_j, C_j^\alpha} \\ \gamma \dot{l}_{N_{j+1}, C_j} = -\kappa l_{N_{j+1}, C_j} + u_{N_{j+1}, C_j} \\ \gamma \dot{l}_{C_{j+1}^\alpha, N_{j+1}} = -\kappa l_{C_{j+1}^\alpha, N_{j+1}} + u_{C_{j+1}^\alpha, N_{j+1}} \end{cases} \quad (4.11)$$

e

$$\begin{cases} \dot{l}_{C_j, C_j^\alpha} = L_{C_j, C_j^\alpha} \\ m\dot{L}_{C_j, C_j^\alpha} = -\gamma L_{C_j, C_j^\alpha} - \kappa l_{C_j, C_j^\alpha} + u_{C_j, C_j^\alpha} \\ \dot{l}_{N_{j+1}, C_j} = L_{N_{j+1}, C_j} \\ m\dot{L}_{N_{j+1}, C_j} = -\gamma L_{N_{j+1}, C_j} - \kappa l_{N_{j+1}, C_j} + u_{N_{j+1}, C_j} \\ \dot{l}_{C_{j+1}^\alpha, N_{j+1}} = L_{C_{j+1}^\alpha, N_{j+1}} \\ m\dot{L}_{C_{j+1}^\alpha, N_{j+1}} = -\gamma L_{C_{j+1}^\alpha, N_{j+1}} - \kappa l_{C_{j+1}^\alpha, N_{j+1}} + u_{C_{j+1}^\alpha, N_{j+1}} \end{cases} \quad (4.12)$$

4.2 Configuração do problema

Após a definição dos modelos dinâmicos, consideramos o seguinte problema de controle:

$$\min_u J[u_{jk}] = \min_u \left(\phi(l_{jk}(t_F)) + \int_{t_0}^{t_F} f(l_{jk}(t), u_{jk}(t)) dt \right) \quad (4.13)$$

onde

$$k = C_j, C_j^\alpha, N_{j+1}, C_j, C_{j+1}^\alpha, N_{j+1} \quad \forall j,$$

e a função controle u é formada por funções que descrevem melhor a dinâmica (os termos de campos de força usuais como, por exemplo, pontes de hidrogênio, ângulos

diedrais e o potencial de Lennard-Jones).

Supomos que f seja uma função homogênea de grau p na variável de estado l e de grau q na variável de controle u :

$$f(l_{jk}(t), u_{jk}(t)) = \sum_k \sum_{j=1}^{n-1} \left((l_{jk}(t))^p + A_{jk}(u_{jk}(t))^q \right)$$

A função de compensação ϕ (*payoff function* no original) é resultado de um controle externo sobre o sistema para extremizar a função f em um tempo específico do intervalo. Por exemplo, em um modelo de epidemia, queremos ainda extremizar também o número de indivíduos infectados a um determinado tempo de evolução. Aqui, consideramos que este tempo específico seja no tempo final e que a função seja homogênea de grau r em l :

$$\phi(l_{jk}(t_F)) = \sum_k \sum_{j=1}^{n-1} (l_{jk}(t_F))^r$$

com p , q e r inteiros positivos.

Em seguida, supomos que a evolução no tempo dos três diferentes tipos de comportamentos de ligações sejam uniformes, ou seja, independentes de j , o que resulta em:

$$\begin{cases} l_{jk} = l_k \\ u_{jk} = u_k, \quad \forall j \\ A_{jk} = A_k \end{cases}$$

e a funcional (4.13) assume a forma:

$$J[u_k] = (n-1) \sum_k (l_k(t_F))^r + (n-1) \int_{t_0}^{t_F} \sum_k \left((l_k(t))^p + A_k(u_k(t))^q \right) dt \quad (4.14)$$

Os sistemas (4.11) e (4.12) também são simplificados, respectivamente, como:

$$\dot{l}_k = -\frac{\kappa}{\gamma} l_k + \frac{1}{\gamma} u_k \quad (4.15)$$

e

$$\begin{cases} \dot{l}_k = L_k \\ \dot{L}_k = -\frac{\gamma}{m} L_k - \frac{\kappa}{m} l_k + \frac{1}{m} u_k \end{cases} \quad (4.16)$$

Com a funcional objetivo (4.14) e com os sistemas (4.15) e (4.16), podemos

escrever as respectivas funções Hamiltonianas:

$$H_0 = (n-1) \sum_k \left((l_k(t))^p + A_k(u_k(t))^q \right) + \sum_k \lambda_{l_k}(t) \dot{l}_k(t) \quad (4.17)$$

e

$$H_1 = (n-1) \sum_k \left((l_k(t))^p + A_k(u_k(t))^q \right) + \sum_k \left(\lambda_{l_k}(t) \dot{l}_k(t) + \lambda_{L_k}(t) \dot{L}_k(t) \right) \quad (4.18)$$

onde $\lambda_k(t)$ e $\mu_k(t)$ são as funções adjuntas do problema.

4.2.1 Caso de termos inerciais desprezíveis diante do termo de fricção

Uma vez que temos (4.14) e (4.15), formamos o Hamiltoniano (4.17)

$$H_0 = (n-1) \sum_k \left((l_k(t))^p + A_k(u_k(t))^q \right) + \sum_k \lambda_{l_k}(t) \left(-\frac{\kappa}{\gamma} l_k(t) + \frac{1}{\gamma} u_k(t) \right)$$

Pela definição de *condição de otimalidade*, temos:

$$\begin{aligned} 0 &= \frac{\partial H_0}{\partial u_k(t)} = (n-1)qA_k(u_k(t))^{q-1} + \frac{\lambda_{l_k}(t)}{\gamma} \\ \lambda_{l_k}(t) &= -(n-1)q\gamma A_k(u_k(t))^{q-1} \end{aligned} \quad (4.19)$$

e a *equação adjunta* é dada por:

$$\dot{\lambda}_{l_k}(t) = -\frac{\partial H_0}{\partial l_k(t)} = -(n-1)p(l_k(t))^{p-1} + \frac{\kappa}{\gamma} \lambda_{l_k}(t) \quad (4.20)$$

A *condição de transversalidade*, por sua vez, é:

$$\lambda_{l_k}(t_F) = \frac{\partial \phi}{\partial l_k(t)}(l_k(t_F)) = (n-1)r(l_k(t_F))^{r-1} \quad (4.21)$$

Substituímos a equação (4.19) juntamente com sua derivada no tempo em (4.20), obtendo a equação para a variável de controle:

$$(q-1)\dot{u}_k(t) = \frac{\kappa}{\gamma} u_k(t) + \frac{p}{q} \frac{1}{\gamma A_k} (l_k(t))^{p-1} (u_k(t))^{2-q} \quad (4.22)$$

Da mesma forma, a condição de transversalidade reescrita em função da variável de controle resulta em:

$$u_k(t_F) = \left(\frac{-r}{q\gamma A_k} \right)^{\frac{1}{q-1}} (l_k(t_F))^{\frac{r-1}{q-1}} \quad (4.23)$$

Unindo (4.15) à (4.22), temos o sistema de equações dinâmicas:

$$\begin{cases} \dot{l}_k = -\frac{\kappa}{\gamma}l_k + \frac{1}{\gamma}u_k \\ (q-1)\dot{u}_k = \frac{\kappa}{\gamma}u_k + \frac{p}{q}\frac{1}{\gamma A_k}l_k^{p-1}u_k^{2-q} \end{cases} \quad (4.24)$$

4.2.2 Caso de pequena diferença entre as massas dos sítios atômicos

O Hamiltoniano (4.18) pode ser expresso de forma explícita, substituindo (4.16):

$$H_1 = (n-1) \sum_k \left((l_k(t))^p + A_k (u_k(t))^q \right) + \sum_k \lambda_{l_k}(t) L_k(t) + \sum_k \lambda_{L_k}(t) \left(-\frac{\gamma}{m} L_k - \frac{\kappa}{m} l_k + \frac{1}{m} u_k \right)$$

A condição de otimalidade pode então ser obtida:

$$\begin{aligned} 0 &= \frac{\partial H_1}{\partial u_k(t)} = (n-1)qA_k(u_k(t))^{q-1} + \frac{\lambda_{L_k}(t)}{m} \\ \lambda_{L_k}(t) &= -(n-1)qmA_k(u_k(t))^{q-1} \end{aligned} \quad (4.25)$$

Nesta configuração do problema, temos duas equações adjuntas:

$$\dot{\lambda}_{l_k}(t) = -\frac{\partial H_1}{\partial l_k(t)} = -(n-1)p(l_k(t))^{p-1} + \frac{\kappa}{m}\lambda_{L_k}(t) \quad (4.26)$$

e

$$\dot{\lambda}_{L_k}(t) = -\frac{\partial H_1}{\partial L_k(t)} = -\lambda_{l_k}(t) + \frac{\gamma}{m}\lambda_{L_k}(t) \quad (4.27)$$

Da mesma forma, as condições de transversalidade são:

$$\lambda_{l_k}(t_F) = \frac{\partial \phi}{\partial l_k(t)}(l_k(t_F)) = (n-1)r(l_k(t_F))^{r-1} \quad (4.28)$$

e

$$\lambda_{L_k}(t_F) = \frac{\partial \phi}{\partial L_k(t)}(L_k(t_F)) = 0 \quad (4.29)$$

Utilizando a equação (4.25), podemos reescrever as equações (4.26) e (4.27) em função da variável de controle reespectivamente como:

$$\dot{\lambda}_{l_k}(t) = -(n-1)p(l_k(t))^{p-1} - (n-1)qA_k\kappa(u_k(t))^{q-1} \quad (4.30)$$

e

$$(q-1)\dot{u}_k(t) = \frac{\gamma}{m}u_k(t) + \frac{1}{(n-1)qmA_k}\lambda_{l_k}(t)(u_k(t))^{2-q} \quad (4.31)$$

Juntamente com a equação (4.29), temos que:

$$u_k(t_F) = 0 \quad (4.32)$$

Da união de (4.16) com (4.30) e (4.31), obtemos o sistema de equações:

$$\begin{cases} \dot{l}_k = L_k \\ \dot{L}_k = -\frac{\gamma}{m}L_k - \frac{\kappa}{m}l_k + \frac{1}{m}u_k \\ \dot{\lambda}_{l_k} = -(n-1)\left(pl_k^{p-1} + qA_k\kappa u_k^{q-1}\right) \\ (q-1)\dot{u}_k = \frac{\gamma}{m}u_k + \frac{1}{(n-1)qmA_k}\lambda_{l_k}u_k^{2-q} \end{cases} \quad (4.33)$$

4.3 Equações Adimensionais

Através de um processo de adimensionalização, selecionamos termos intrínsecos ao sistema de forma que o produto com as constantes dimensionais resultem em fatores adimensionais, e o conjunto de equações seja expresso por variáveis adimensionais. Com isto, o número de parâmetros, agora reunidos em grupos sem dimensão, é reduzido em relação ao sistema original [21].

O número máximo de parâmetros adimensionais a serem incluídos ficam restritos à diferença entre a quantidade de equações de condição e a de variáveis do sistema, incluindo a variável independente. Para diferentes escolhas destes parâmetros, diferentes características do sistema de equações podem ser evidenciadas [22].

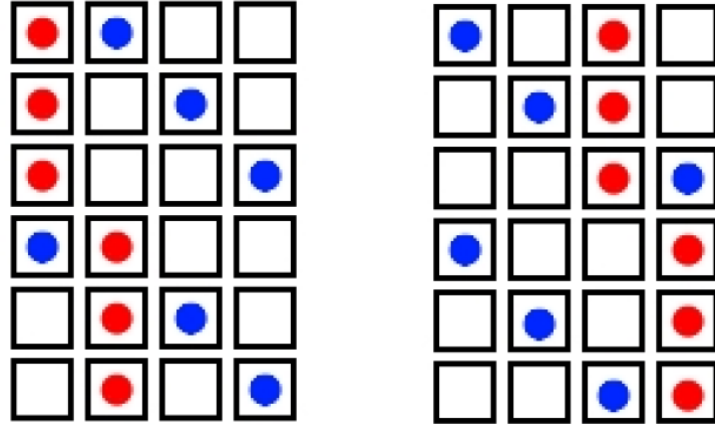


Figura 4.2: Exemplo de combinações com quatro caixas distintas, uma bola azul e uma bola vermelha.

Fazemos a seguinte comparação para demonstrar a vasta gama de possibilidades possíveis na escolha da inclusão de parâmetros adimensionais:

Consideremos que as equações de condição sejam como caixas distintas e os

parâmetros adimensionais que podem ser incluídos sejam bolas também distintas entre si. Cada caixa só pode receber uma bola e podemos escolher se utilizaremos todas as bolas ou não. Seja, por exemplo, um caso em que temos quatro caixas distintas e duas bolas, uma vermelha e uma azul (Figura 4.2). O número de combinações possíveis utilizando as duas bolas é dado por:

$$C_4^1 \cdot C_{4-1}^1 = \frac{4!}{1!(4-1)!} \cdot \frac{(4-1)!}{1!(4-1-1)!} = \frac{4!}{1!3!} \cdot \frac{3!}{1!2!} = \frac{4!}{1!1!2!} = 4 \cdot 3 = 12 .$$

Temos, então, no máximo 12 combinações possíveis. Se for escolhido utilizar apenas uma bola, teríamos quatro combinações possíveis, enquanto que se houvesse mais uma ou duas bolas, teríamos 24 combinações (o número máximo de bolas não pode ultrapassar o número de caixas). Logo, para n caixas distintas e m tipos de bola, teríamos:

$$C_n^1 \cdot C_{n-1}^1 \cdots C_{n-m+1}^1 = \frac{n!}{1!(n-1)!} \cdot \frac{(n-1)!}{1!(n-2)!} \cdots \frac{(n-m+1)!}{1!(n-m)!} = \frac{n!}{1! \dots 1!(n-m)!}$$

onde $m \leq n$.

Para casos em que há um grande número de equações de condição e também um grande número de parâmetros adimensionais que possam ser incluídos, o número de combinações possíveis é muito grande para que todas as escolhas sejam vasculhadas.

4.3.1 Caso de termos inerciais desprezíveis diante do termo de fricção

Para o sistema (4.24), introduzimos variáveis adimensionais a partir das variáveis dependentes l e u e da variável independente t :

$$\begin{aligned} t^* &= \frac{t}{\alpha_t} , \\ l^* &= \frac{l}{\alpha_l} , \\ u^* &= \frac{u}{\alpha_u} \end{aligned}$$

onde (*) denota que a variável é adimensional. Os parâmetros α_t , α_l e α_u têm, respectivamente, dimensões de tempo (T), comprimento (L) e força (LMT^{-2}). Re-escrevendo o sistema, temos:

$$\begin{cases} \frac{\alpha_l}{\alpha_t} \dot{l}_k^* = -\frac{\kappa}{\gamma} \alpha_l l_k^* + \frac{1}{\gamma} \alpha_u u_k^* \\ (q-1) \frac{\alpha_u}{\alpha_t} \dot{u}_k^* = \frac{\kappa}{\gamma} \alpha_u u_k^* + \frac{p}{q} \frac{1}{\gamma A_k} \alpha_l^{p-1} \alpha_u^{2-q} (l_k^*)^{p-1} (u_k^*)^{2-q} \end{cases}$$

ou

$$\begin{cases} \dot{l}_k^* = -\frac{\kappa}{\gamma}\alpha_t l_k^* + \frac{1}{\gamma}\frac{\alpha_u\alpha_t}{\alpha_l}u_k^* \\ (q-1)\dot{u}_k^* = \frac{\kappa}{\gamma}\alpha_t u_k^* + \frac{p}{q}\frac{1}{\gamma A_k}\alpha_t\alpha_l^{p-1}\alpha_u^{1-q}(l_k^*)^{p-1}(u_k^*)^{2-q} \end{cases}$$

O sistema contém três variáveis, sendo um independente (t) e duas dependentes (l e u) e três equações de condição. Portanto, nenhum parâmetro adimensional pode ser incluído. A relação entre os parâmetros dimensionais é dada então por:

$$\begin{aligned} 1 &= \frac{\kappa}{\gamma}\alpha_t, \\ 1 &= \frac{1}{\gamma}\frac{\alpha_u\alpha_t}{\alpha_l}, \\ 1 &= \frac{1}{\gamma A}\alpha_t\alpha_l^{p-1}\alpha_u^{1-q} \end{aligned}$$

Após as devidas substituições e manipulações, temos que:

$$\begin{aligned} \alpha_t &= \frac{\gamma}{\kappa}, \\ \alpha_l^{p-q} &= A\kappa^q, \\ \alpha_u^{p-q} &= A\kappa^p \end{aligned}$$

O sistema completo adimensional (suprimindo os asteriscos) apresenta a seguinte forma:

$$\begin{cases} \dot{l}_k = -l_k + u_k \\ (q-1)\dot{u}_k = u_k + \frac{p}{q}l_k^{p-1}u_k^{2-q} \end{cases} \quad (4.34)$$

4.3.2 Caso de pequena diferença entre as massas dos sítios atômicos

De forma semelhante, introduzimos variáveis adimensionais para o sistema (4.33), com adição das variáveis dependentes L e λ :

$$\begin{aligned} t^* &= \frac{t}{\beta_t}, \\ l^* &= \frac{l}{\beta_l}, \\ u^* &= \frac{u}{\beta_u}, \\ L^* &= \frac{L}{\beta_L}, \\ \lambda_l^* &= \frac{\lambda_l}{\beta_{\lambda_l}} \end{aligned}$$

Os parâmetros β_t , β_l , β_L , β_{λ_l} e β_u têm, respectivamente, dimensões T , L , LT^{-1} , $L^{p-1}T$ e LMT^{-2} . O sistema pode ser reescrito como:

$$\begin{cases} \frac{\beta_l}{\beta_t} \dot{l}_k^* = \beta_L L_k^* \\ \frac{\beta_L}{\beta_t} \dot{L}_k^* = -\frac{\gamma}{m} \beta_L L_k^* - \frac{\kappa}{m} \beta_l l_k^* + \frac{1}{m} \beta_u u_k^* \\ \frac{\beta_{\lambda_l}}{\beta_t} \dot{\lambda}_{l_k}^* = -(n-1) \left(p \beta_l^{p-1} (l_k^*)^{p-1} + q A \kappa \beta_u^{q-1} (u_k^*)^{q-1} \right) \\ (q-1) \frac{\beta_u}{\beta_t} \dot{u}_k^* = \frac{\gamma}{m} \beta_u u_k^* + \frac{1}{(n-1)q} \frac{1}{mA} \frac{\beta_{\lambda_l}}{\beta_u^{q-2}} \lambda_{l_k}^* (u_k^*)^{2-q} \end{cases}$$

ou

$$\begin{cases} \dot{l}_k^* = \frac{\beta_L \beta_t}{\beta_l} L_k^* \\ \dot{L}_k^* = -\frac{\gamma}{m} \beta_t L_k^* - \frac{\kappa}{m} \frac{\beta_l \beta_t}{\beta_L} l_k^* + \frac{1}{m} \frac{\beta_u \beta_t}{\beta_L} u_k^* \\ \dot{\lambda}_{l_k}^* = -(n-1) \left(p \frac{\beta_l^{p-1} \beta_t}{\beta_{\lambda_l}} (l_k^*)^{p-1} + q A \kappa \frac{\beta_u^{q-1} \beta_t}{\beta_{\lambda_l}} (u_k^*)^{q-1} \right) \\ (q-1) \dot{u}_k^* = \frac{\gamma}{m} \beta_t u_k^* + \frac{1}{(n-1)q} \frac{1}{mA} \frac{\beta_{\lambda_l} \beta_t}{\beta_u^{q-1}} \lambda_{l_k}^* (u_k^*)^{2-q} \end{cases}$$

Uma vez que o sistema tem cinco variáveis (uma independente e quatro dependentes) e sete equações de condição, podemos incluir no máximo dois parâmetros adimensionais, o que nos dá 42 possibilidades ($C_7^1 \cdot C_6^1$). A inclusão dos parâmetros é feita da seguinte forma:

$$\begin{aligned} 1 &= \frac{\beta_L \beta_t}{\beta_l}, \\ a &= \frac{\gamma}{m} \beta_t, \\ b &= \frac{\kappa}{m} \frac{\beta_l \beta_t}{\beta_L}, \\ 1 &= \frac{1}{m} \frac{\beta_u \beta_t}{\beta_L}, \\ 1 &= \frac{\beta_l^{p-1} \beta_t}{\beta_{\lambda_l}}, \\ 1 &= A \kappa \frac{\beta_u^{q-1} \beta_t}{\beta_{\lambda_l}}, \\ 1 &= \frac{1}{mA} \frac{\beta_{\lambda_l} \beta_t}{\beta_u^{q-1}} \end{aligned}$$

onde a e b são os parâmetros introduzidos. Os cálculos para a determinação dos parâmetros adimensionais em relação às constantes dimensionais do sistema original,

resultam em:

$$a = \frac{\gamma}{\sqrt{\kappa m}}$$

e

$$b = 1$$

Após as substituições e manipulações, temos os parâmetros dimensionais:

$$\begin{aligned}\beta_t &= \sqrt{\frac{m}{\kappa}}, \\ \beta_l^{p-q} &= A\kappa^q, \\ \beta_u^{p-q} &= A\kappa^p, \\ \beta_L^{p-q} &= A\kappa^q \left(\frac{\gamma}{ma}\right)^{p-q}, \\ \beta_{\lambda_l}^{p-q} &= A^{p-1} \kappa^{q(p-1)} \left(\frac{ma}{\gamma}\right)^{p-q}\end{aligned}$$

E o sistema adimensional completo (suprimido o asterisco) é reescrito como:

$$\begin{cases} \dot{l}_k = L_k \\ \dot{L}_k = -aL_k - l_k + u_k \\ \dot{\lambda}_{l_k} = -(n-1)(pl_k^{p-1} + qu_k^{q-1}) \\ (q-1)\dot{u}_k = au_k + \frac{1}{(n-1)q}\lambda_{l_k}u_k^{2-q} \end{cases} \quad (4.35)$$

Capítulo 5

Resultados

Com o auxílio do software Maple, geramos gráficos dos estados l_k (comprimentos de ligação dos átomos da espinha dorsal do polipeptídeo) e dos controles u_k . Tanto para o caso de massas atômicas desprezíveis quanto para o caso de pequena diferença entre as massas atômicas, fazemos a análise para os seguintes valores:

- $p = 1, q = 2$ (estado l_1 e controle u_1);
- $p = 2, q = 2$ (estado l_2 e controle u_2).

Essa escolha se deve ao fato das soluções analíticas nestes casos serem facilmente obtidas. Em um segundo momento, a utilização de outros conjuntos de valores (com $p \geq 1$ e $q \geq 2$) pode mostrar resultados interessantes, mas deve ser feita por uma aproximação de solução numérica.

5.1 Caso de termos inerciais desprezíveis diante do termo de fricção

Para a condição (4.23), considerando $A > 0$ e r um número inteiro positivo, temos:

$$u_k(t_F) < 0$$

Como l é uma medida de comprimento, não pode apresentar valores negativos durante o intervalo ΔT . Uma vez que temos estas restrições, analisamos o sistema (4.34) e determinamos valores limites para t_F .

As condições iniciais utilizadas são:

$$l_1(t_0) = 1, l_2(t_0) = 1, u_1(t_0) = -1, u_2(t_0) = -1$$

Na Figura 5.1 temos os estados l_1 e l_2 do sistema de massas atômicas desprezíveis,

e, como pode ser verificado na Figura 5.2 os controles u_1 e u_2 são negativos durante todo o intervalo ΔT .

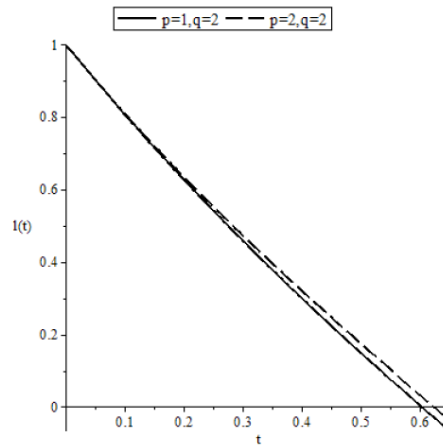


Figura 5.1: Comparação dos estados com termos inerciais desprezíveis diante do termo de fricção, para os casos $p = 1, q = 2$ e $p = 2, q = 2$.

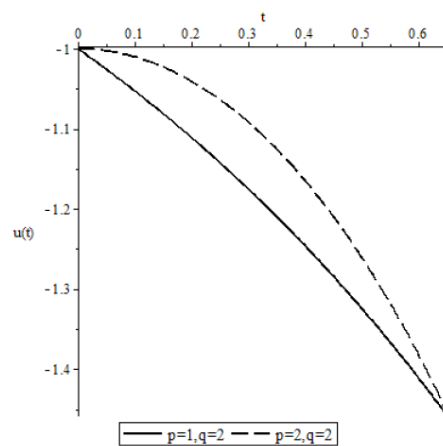


Figura 5.2: Comparação dos controles com termos inerciais desprezíveis diante do termo de fricção, para os casos $p = 1, q = 2$ e $p = 2, q = 2$.

Os valores máximos de t_F encontrados foram:

$$t_{1F} \approx 0.603$$

$$t_{2F} \approx 0.623$$

e podem ser verificados na Figura 5.3 e na Figura 5.4, respectivamente.

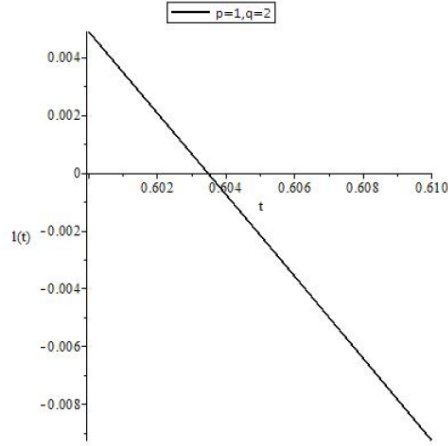


Figura 5.3: t_F máximo para o caso $p = 1, q = 2$. Com a restrição $l_1 \geq 0$ durante o intervalo ΔT , temos que $t_{1F} = 0.603$.

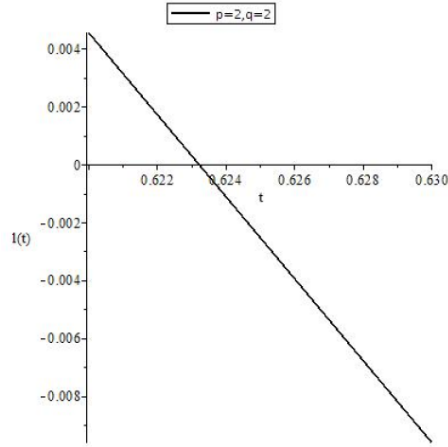


Figura 5.4: t_F máximo para o caso $p = 2, q = 2$. Com a restrição $l_2 \geq 0$ durante o intervalo ΔT , temos que $t_{2F} = 0.623$.

5.2 Caso de pequena diferença entre as massas dos sítios atômicos

Para o sistema (4.35), alguns valores do parâmetro adimensional a incluído são testados. A condição (4.32) determina t_F e, juntamente com a restrição a valores negativos de l , indica os valores mínimos para a .

As condições iniciais utilizadas foram:

$$l_1(t_0) = 1, L_1(t_0) = -1.5, \lambda_1(t_0) = 0, u_1(t_0) = 1$$

$$l_2(t_0) = 1, L_2(t_0) = -2, \lambda_2(t_0) = 0, u_2(t_0) = 1$$

Para $a = 1$, a Figura 5.5 e a Figura 5.6 mostram, respectivamente, os estados l_1 e l_2 e os controles u_1 e u_2 .

A Figura 5.7 e a Figura 5.8 mostram os valores de t_F encontrados.

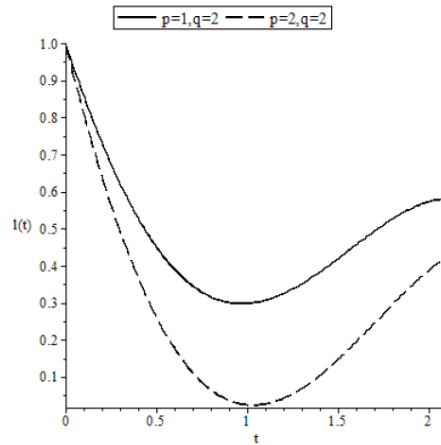


Figura 5.5: Comparação dos estados com $\Delta m \approx 0$ e $a = 1$ para os casos $p = 1, q = 2$ e $p = 2, q = 2$.

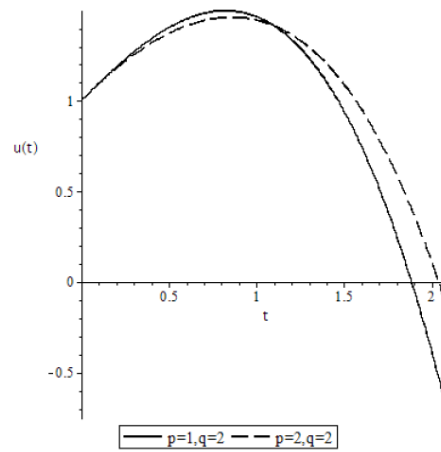


Figura 5.6: Comparação dos controles com $\Delta m \approx 0$ e $a = 1$ para os casos $p = 1, q = 2$ e $p = 2, q = 2$.

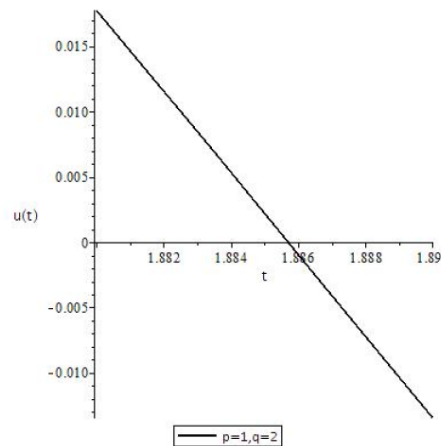


Figura 5.7: t_F para o caso $p = 1, q = 2$. O tempo final é determinado pela condição $u(t_F) = 0$. Aqui, $t_{1F} \approx 1.886$

Os valores foram, respectivamente:

$$t_{1F} \approx 1.886$$

$$t_{2F} \approx 2.040$$

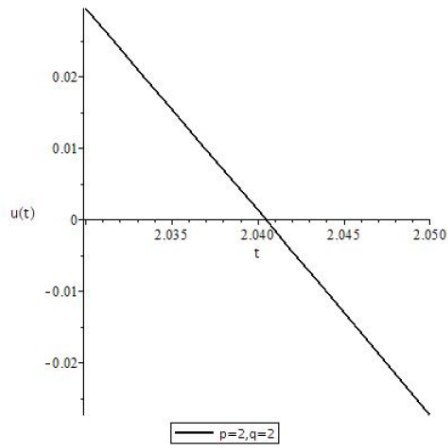


Figura 5.8: t_F para o caso $p = 2, q = 2$. O tempo final é determinado pela condição $u(t_F) = 0$. Aqui, $t_{2F} \approx 2.040$

A Figura 5.9 mostra o estado l_1 e a Figura 5.10 o controle u_1 com $a = 0.597$, o valor mínimo para o caso $p = 1, q = 2$.

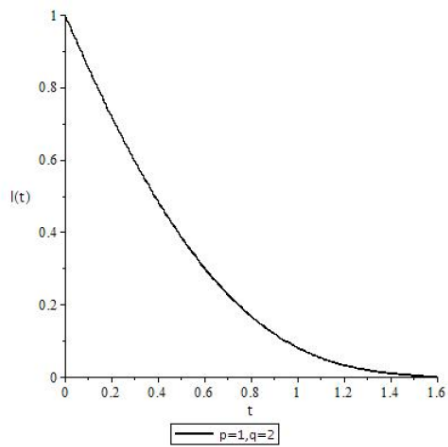


Figura 5.9: Gráfico do estado l_1 para o valor do parâmetro adimensional $a = 0.597$.

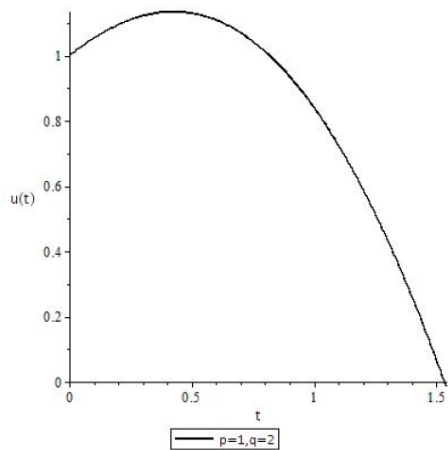


Figura 5.10: Gráfico do controle u_1 para o valor do parâmetro adimensional $a = 0.597$.

O tempo final obtido foi:

$$t_{1F} \approx 1.533$$

como pode ser verificado na Figura 5.11. A Figura 5.12, por sua vez, garante que o estado l_1 não cruza o eixo antes de t_{1F} .

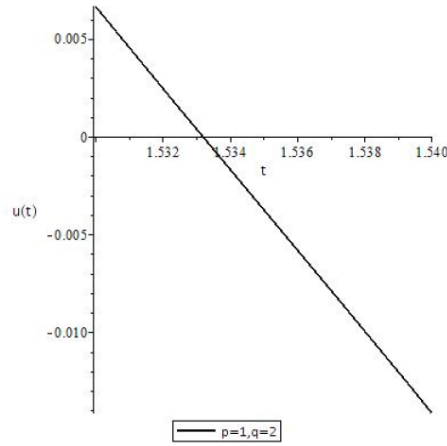


Figura 5.11: t_{1F} para $a = 0.597$. Com a condição $u(t_F) = 0$, temos que $t_{1F} \approx 1.533$.

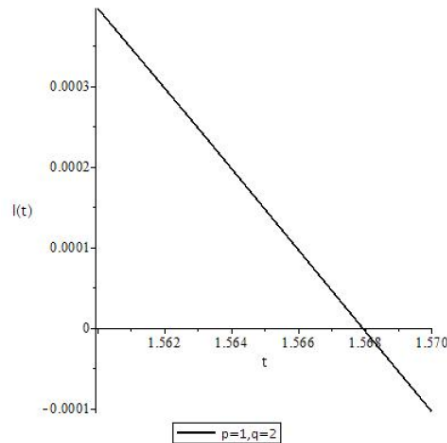


Figura 5.12: Valor de t em que l_1 cruza o eixo para $a = 0.597$.

O estado e o controle obtidos com valor mínimo $a = 0.965$ para $p = 2, q = 2$ podem ser verificados nas Figuras 5.13 e 5.14.

Como mostrado na Figura 5.15, o valor final para o sistema com $p = 2, q = 2$ e $a = 0.965$ é:

$$t_{2F} \approx 2.009$$

Para valores próximos a dois, o estado e o controle assumem valores muito maiores que os obtidos com $a = 1$ e o crescimento de suas dimensões são muito maiores que o crescimento do tempo.

Para $a = 1.937$, o estado l_1 e o controle u_1 obtidos são mostrados, respectivamente, nas Figuras 5.16 e 5.17.

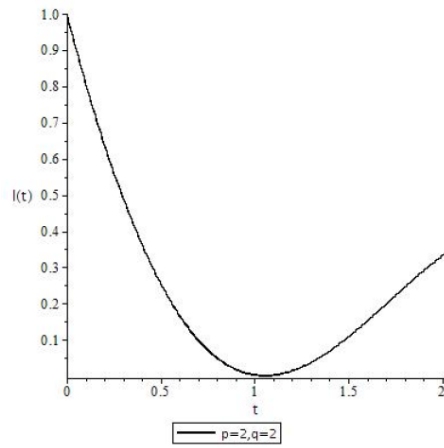


Figura 5.13: Gráfico do estado l_2 para o valor do parâmetro adimensional $a = 0.965$.

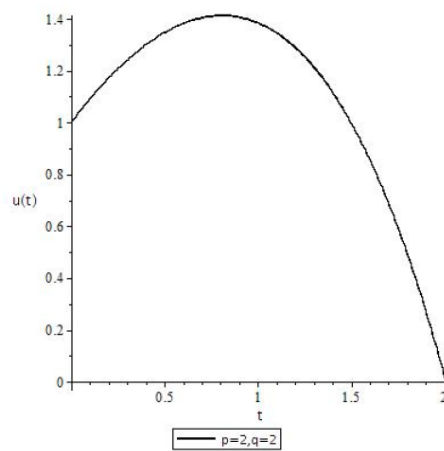


Figura 5.14: Gráfico do controle u_2 para o valor do parâmetro adimensional $a = 0.965$.

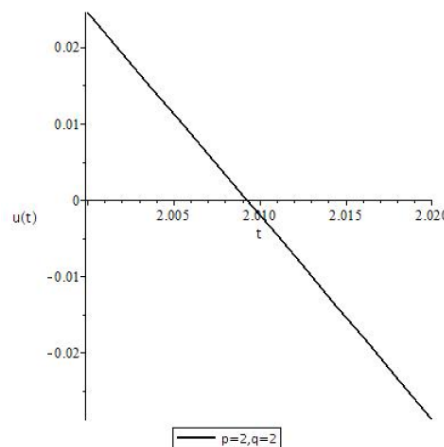


Figura 5.15: t_{2F} para $a = 0.965$. Com a condição $u(t_F) = 0$, temos que $t_{2F} \approx 2.009$.

Na Figura 5.18 temos o valor do tempo final:

$$t_{1F} \approx 9.978$$

Da mesma forma que para o caso $p = 1, q = 2$, conforme a aumenta, os valores

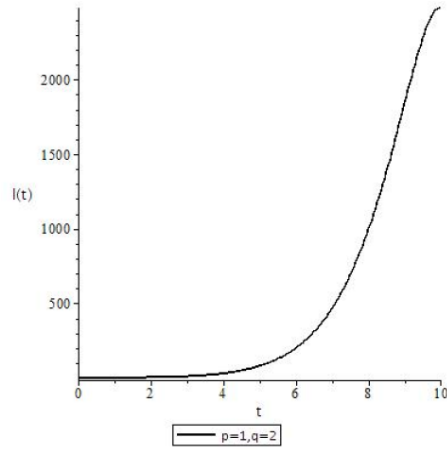


Figura 5.16: Gráfico do estado l_1 para o valor do parâmetro adimensional $a = 1.937$.

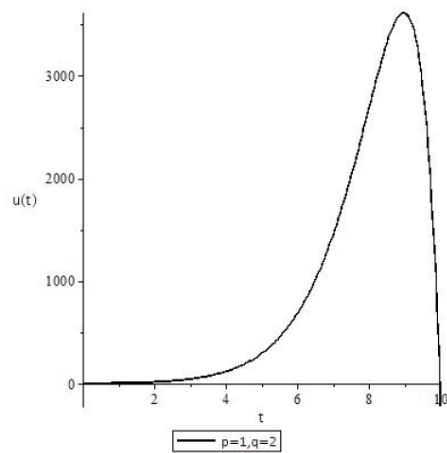


Figura 5.17: Gráfico do controle u_1 para o valor do parâmetro adimensional $a = 1.937$.

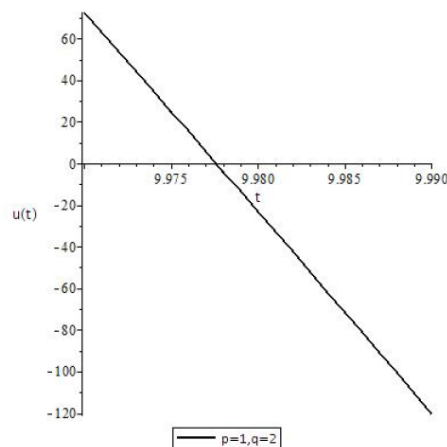


Figura 5.18: t_{1F} para $a = 1.937$. Com a condição $u(t_F) = 0$, temos que $t_{1F} \approx 9.978$.

do estado e do controle têm o crescimento de suas dimensões muito maior que o crescimento do tempo. O estado l_2 e o controle u_2 são mostrados nas Figuras 5.19 e 5.20 para $a = 2.094$.

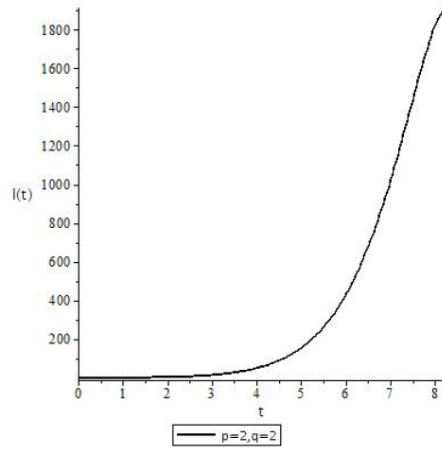


Figura 5.19: Gráfico do estado l_2 para o valor do parâmetro adimensional $a = 2.094$.

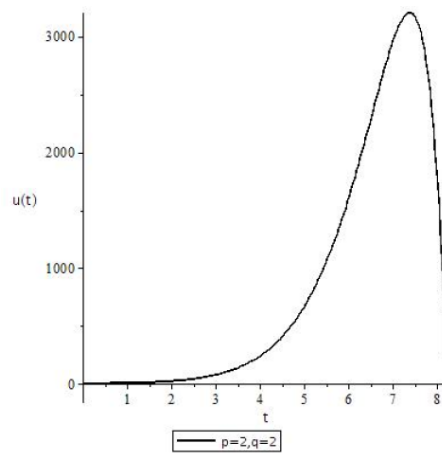


Figura 5.20: Gráfico do controle u_2 para o valor do parâmetro adimensional $a = 2.094$.

O valor final do tempo aqui obtido, como pode ser verificado na Figura 5.21 é:

$$t_{2F} \approx 8.235$$

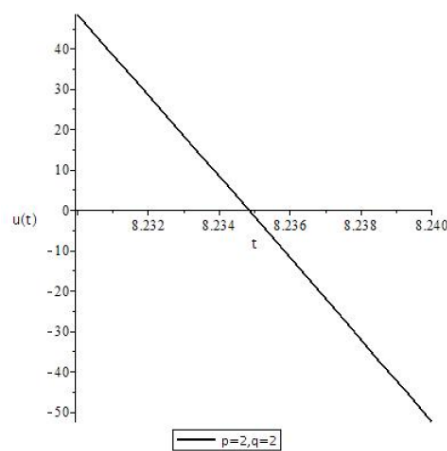


Figura 5.21: t_{2F} para $a = 2.094$. Com a condição $u(t_F) = 0$, temos que $t_{2F} \approx 8.235$.

5.3 Discussão sobre os resultados

Os dois casos analisados tiveram resultados bem distintos. Para o caso em que os termos inerciais são considerados muito pequenos diante do termo de fricção, o sistema é muito mais simples em relação ao outro, pois como não podemos incluir um parâmetro adimensional, ficamos restritos a analisar os comportamentos do estado e do controle de forma a determinar o intervalo de tempo em que nenhuma condição é violada. Ambas são monótonas durante todo o intervalo e o tempo final deve ser tal que o estado não chegue a zero.

O caso de pequena diferença entre as massas dos sítios atômicos, por sua vez, com a inclusão do parâmetro adimensional a , mostrou resultados mais interessantes. O tratamento do problema como pequenas vibrações fica mais evidente com comportamento do estado neste caso. O tempo final ocorre quando o controle chega a zero e o valor mínimo de a é limitado pela condição do estado ser positivo durante todo o intervalo ΔT . Na Tabela 5.1 são mostrados os tempos finais t_F obtidos para os diferentes valores de a , com $p = 1, q = 2$ e $p = 2, q = 2$.

Tabela 5.1: O comportamento detectado nos dois casos, $p = 1, q = 2$ e $p = 2, q = 2$ é que para $a < 1.000$, quando $u(t_F) = 0$, $l(t_F) \geq 0$ aí. Para $a > 1.000$, quando $u(t_F) = 0$, $l(t_F) \gg 1$.

	a	t_F
$p = 1, q = 2$	0.597	1.533
	1.000	1.886
	1.937	9.978
$p = 2, q = 2$	0.965	2.009
	1.000	2.040
	2.094	8.235

Capítulo 6

Conclusão

Neste trabalho, foi suposto o caso em que uma molécula de proteína já se encontra enovelada em um mínimo de energia e seu comportamento é estudado através de pequenas vibrações. O modelo adotado considera os sítios atômicos dos elementos presentes na espinha dorsal do peptídeo e não apenas dos C^α 's. O problema é tratado com métodos de Pontryagin e valores para as potências p e q são atribuídos apenas na análise e não durante a formulação para obter métodos de forma quadrática.

A comparação feita entre o caso em que os termos inerciais são considerados desprezíveis diante do termo de fricção e o caso de diferença entre as massas aproximadamente zero, indicou que o último retrata melhor o sistema. A esta aproximação, após o processo de adimensionalização, foi considerado o modelo com um parâmetro adimensional a , e assim pudemos observar a resposta do sistema para diferentes valores adotados.

Como trabalhos futuros, algumas alterações são previstas, como:

- Utilização de controles limitados;
- Consideração de controle com variação do tempo final, o que leva a $H(t_F) = 0$;
- Alterações no modelo dinâmico, levando em conta a estrutura da molécula de proteína na escrita de uma lagrangiana efetiva para pequenas oscilações.

Referências Bibliográficas

- [1] HUANG, K. *Lectures on Statistical Physics and Protein Folding*. World Scientific, 2005.
- [2] MONDAINI, R. P., DE ALBUQUERQUE NETO, S. C. “BIOMAT 2014 International Symposium on Mathematical and Computational Biology”. cap. Optimal Control of a Coarse-Grained Model for Protein Dynamics, pp. 12–25, World Scientific Co. Pte. Ltd., 2015.
- [3] ALBERTS, B., JOHNSON, A., LEWIS, J., et al. *Molecular Biology of The Cell*. Garland Science, 2008.
- [4] FERSHT, A. *Structure and Mechanism in Protein Folding*. W. H. Freeman and Company, 1999.
- [5] NEUMAIER, A. “Molecular Modelling of Proteins and Mathematical Prediction of Protein Structures”, *SIAM Rev.*, v. 39, pp. 407–460, 1997.
- [6] HUANG, E. S., SUBBIAH, S., LEVITT, M. “Recognizing Native Folds by the Arrangement of Hydrophobic and Polar Residues”, *Journal of Molecular Biology*, v. 252, pp. 709–720, 1995.
- [7] BEN-NAIM, A. “Levinthal’s Paradox Revisited, and Dismissed”, *Open Journal of Biophysics*, v. 2, pp. 23–32, 2012.
- [8] AMADEI, A., DE GROOT, B. L., CERUSO, M.-A., et al. “A Kinetic Model for the Internal Motions of Proteins: Diffusion Between Multiple Harmonic Wells”, *Proteins: Structure, Function, and Bioinformatics*, v. 35, pp. 283–292, 1999.
- [9] SCHÜTTE, C., WINKELMANN, S., HARTMANN, C. “Optimal control of molecular dynamics using Markov state models”, *Mathematical Programming (Series B)*, v. 134, pp. 259–282, 2012.
- [10] ACBAS, G., NIESSEN, K. A., SNELL, E. H., et al. “Optical measurements of long-range protein vibrations”, *Nature Communications*, v. 5, 2014. doi: 10.1038/ncomms4076.

- [11] TURTON, D. A., SENN, H. M., HARWOOD, T., et al. “Terahertz underdamped vibrational motion governs protein-ligand and binding in solution”, *Nature Communications*, v. 5, 2014. doi: 10.1038/ncomms4999.
- [12] VERETNIK, S., GU, J., WODAK, S. “Structural Bioinformatics”. 2 ed., cap. Identifying Structural Domains in Proteins, pp. 487–515, Wiley-Blackwell, 2009.
- [13] ERMAN, B., DILL, K. “Gaussian model of protein folding”, *Journal of Chemical Physics*, v. 112, n. 2, pp. 1050–1056, 2000.
- [14] HOANG, T. X., CIEPAK, M. “Molecular dynamics of folding of secondary structures in Go-type models of proteins”, *Journal of Chemical Physics*, v. 112, n. 15, pp. 6851–6862, 2000a.
- [15] TAYLOR, W. R. “Protein structural domain identification”, *Protein Engineering Design & Selection*, v. 12, n. 3, pp. 203–216, 1999.
- [16] TAKADA, S. “Go-ing for the prediction of protein folding mechanisms”, *PNAS*, v. 96, n. 21, pp. 11698–11700, 1999.
- [17] LIN, M. M., ZEWAİL, A. H. “Protein folding - simplicity in complexity”, *Annalen der Physik*, v. 524, pp. 379–391, 2012.
- [18] DINNER, A. R., SALI, A., SMITH, L. J., et al. “Understanding protein folding via free-energy surfaces from theory and experiment”, *Trends in Biochemical Sciences*, v. 25, pp. 331–339, 2000.
- [19] ARKUN, Y., GUR, M. “Combining Optimal Control Theory and Molecular Dynamics for Protein Folding”, *PLoS ONE*, v. 7, pp. e29628, 2012. doi: 10.1371/journal.pone.0029628.
- [20] CIEPAK, M., HOANG, T. X. “Universality Classes in Folding Times of Proteins”, *Biophysical Journal*, v. 84, pp. 475–488, 2003.
- [21] SEGEL, L. “Simplification and Scaling”, *SIAM*, v. 14, n. 4, pp. 547–571, 1972.
- [22] MURRAY, J. D. *Mathematical Biology. I. An introduction*. Springer, 2002.