



Universidade Federal do Triângulo Mineiro
Programa de Pós-Graduação Multicêntrico em
Química de Minas Gerais



UELITOM GALDINO BORGES

ESTUDO DA TERMOESTABILIDADE DE PROTEÍNAS COLD SHOCK
HOMÓLOGAS POR MODELOS TEÓRICOS SIMPLIFICADOS.

Uberaba

2018

Uelitom Galdino Borges

Estudo da termoestabilidade de proteínas cold shock
homólogas por modelos teóricos simplificados.

Dissertação apresentada ao Programa de Pós-Graduação Multicêntrico em Química de Minas Gerais, área de concentração Química Teórica, da Universidade do Triângulo Mineiro, como requisitos necessários à obtenção do título de Mestre em Química.

Orientador: Prof. Dr. Ronaldo Júnio de Oliveira

Uberaba

2018

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

B734e Borges, Uelitom Galdino
 Estudo da termoestabilidade de proteínas cold shock homólogo-
gas por modelos teóricos simplificados / Uelitom Galdino Borges.
-- 2018.
 65 f. : il., fig., graf., tab.

 Dissertação (Mestrado em Química) -- Universidade Federal
do Triângulo Mineiro, Uberaba, MG, 2018
 Orientador: Prof. Dr. Ronaldo Júnio de Oliveira

 1. Proteínas. 2. Temperatura - Estabilidade. 3. Dinâmica mole-
cular. I. Oliveira, Ronaldo Júnio de. II. Universidade Federal do
Triângulo Mineiro. III. Título.

CDU 577.322

UELITOM GALDINO BORGES

Estudo da termoestabilidade de proteínas *Cold Shock* homólogas por modelos teóricos simplificados

Dissertação apresentada ao Curso de Pós-Graduação Multicêntrico em Química de Minas Gerais, área de concentração Química Teórica, da Universidade Federal do Triângulo Mineiro, como requisito parcial para a obtenção do título de mestre em Química.

Uberaba, 18 de Dezembro de 2018.

Banca examinadora:



Prof. Dr. Ronaldo Junio de Oliveira

Universidade Federal do Triângulo Mineiro



Prof. Dr. Francis Barbosa Ferreira

Faculdades Associadas de Uberaba - FAZU



Dr. Gabriel Gouvêa Slade

Pós Doc – Programa de Pós-Graduação Multicêntrico em Química de Minas Gerais

Aos meus amados pais, Iza e Dorisvaldo, pelo apoio incondicional na minha vida.

AGRADECIMENTOS

Agradeço primeiramente ao meu pai e minha mãe, que sempre me deram força para continuar sempre batalhando, e assim conquistando os meus sonhos.

A minha esposa Sueli Noronha Borges e meu filho Samuel Henrique Galdino Borges que sempre me apoiaram a tentar coisas novas e nunca desistir dos meus sonhos, que fizeram deles também esses sonhos. Ao Prof. Dr. Ronaldo Júnio de Oliveira por me receber no seu grupo de estudo e ser meu orientador, bem como os seus orientandos.

Ao Me. Frederico Campos Freitas que com enorme paciência me ensinou os comandos, tanto para a instalação como execução de programas no Linux/Debian.

Aos amigos de Faculdade Lucas Rodrigo Custódio, Roney Melo e Clea de Bosco Rosa Santos e os novos amigos que fui conhecendo durante o mestrado, que compartilharam momentos alegres e difíceis nesses 2 anos de mestrados.

A Universidade Federal do Triângulo Mineiro, aos professores pelo apoio institucional.

A todos o meu respeito e sincero agradecimento, sem vocês esta conquista não seria possível.

“Aqueles que se sentem satisfeitos sentam-se e nada fazem.

Os insatisfeitos são os únicos benfeitores do mundo”.

(Walter S. Landor)

RESUMO

As funções que as proteínas desempenham são de extrema importância para os seres vivos. Essas funções estão ligadas diretamente a sua termoestabilidade, podendo sofrer variações de temperatura. Sendo a termoestabilidade o foco deste trabalho, mais especificamente abordando as diferentes temperaturas de transição entre o enovelamento/desenovelamento das proteínas homologas. Serão utilizadas como objeto de estudo e ponto de partida para essas e outras questões, proteínas *Cold Shock* homologas das bactérias psicrófila, *Listeria monocytogenes* (*Cs-pLa*), mesófila, *Bacillus caldolyticus* (*Bs-CspB*), termófila, *Bacillus subtilis* (*Bc-Csp*) e hipertermófila, *Thermotoga marítima* (*Tm-Csp*), frente as variações de temperatura e de suas interações, levando em conta os aspectos estruturais dessas proteínas bem com a sua identidade e similaridade. Pode-se salientar que existem vários métodos de determinação das estruturas das proteínas tanto em *in vitro* quanto *in silício*. Neste trabalho propõe-se um método *in silício*, no qual aborda a simulação da estrutura da proteína e da otimização, considerando, simultaneamente, os aspectos energéticos e estruturais das proteínas. Para as simulações Dinâmica Molecular utilizou-se o software GROMACS. Para melhor compreender essas diferenças termodinâmicas e a termoestabilidade que ocorre devido as variações das cadeias nas proteínas *Cold Shock* realizamos simulações computacionais levando apenas em consideração o carbono alfa ($C\alpha$) que diminui consideravelmente o tempo de simulação. Os resultados teóricos encontrados corroboraram com os dados experimentais que mostram que a proteína *Cold Shock* bactéria Hipertermófila *Thermotoga marítima* apresenta o maior valor de temperatura de enovelamento e a maior diferença de RMSD entre todas as quatro proteínas *Cold Shock* estudadas. Devido a constatação de que as proteínas *Cold Shock* apresentarem termoestabilidade favorável as alterações de resíduos podem ser realizadas no intuito de serem utilizadas em processos que necessitem alteração na temperatura.

Palavras-chave: Cold Shock. Proteínas homologas. Termoestabilidade. Dinâmica Molecular. Modelo baseado em estrutura. Superfície de energia.

ABSTRACT

The functions that proteins perform are extremely important to living things, this function is directly linked to its thermostability, and may undergo temperature variations. As thermostability is the focus of this work, specifically addressing the different transition temperatures between the unfolding / unfolding of the homologous proteins, using as study object and starting point for these and other questions, Cold Shock homologous proteins of the psychophilic bacteria, *Listeria monocytogenes* (Cs-pLa), mesophilic, *Bacillus caldolyticus* (Bs-CspB), thermophilic, *Bacillus subtilis* (Bc-Csp) and hyperthermophilic, *Thermotoga maritime* (Tm-Csp), considering the variations in temperature and their interactions, taking into account the aspects of these proteins as well as their identity and similarity. It may be noted that there are several methods of determining protein structures in both in vitro and in silicon. In this work, we propose an in silico method, which discusses the simulation of protein structure and optimization, considering simultaneously the energetic and structural aspects of proteins. For the Molecular Dynamics simulations the GROMACS software was used. To better understand these thermodynamic differences and the thermostability that occurs due to the variations of the chains in Cold Shock proteins. We performed computational simulations taking into account only the alpha carbon (C α) that considerably reduces the simulation time. The theoretical results found corroborated with experimental data showing that the Cold Shock protein Hyperthermophilic *Thermotoga marine* bacteria present the highest value of folding temperature and the largest difference of RMSD among all four Cold Shock proteins currently studied. Due to the finding that the Cold Shock proteins presented favorable thermostability to the residue changes that can be carried out in order to be used in processes that require a change in temperature.

Keywords: Cold Shock. Homologous proteins. Thermostability. Molecular dynamics. Structure-based model. Energy Landscape.

LISTA DE FIGURAS

Figura 1 – Estrutura quaternária da mioglobina, primeira proteína a ter sua estrutura determinada por raio-X.	18
Figura 2 – Estrutura típica de um aminoácido, formado por um grupo amina, um grupo carboxila e um radical.	18
Figura 3 – Estrutura de um alfa-aminoácidos na forma de íon dipolar.	19
Figura 4 – Os 20 tipos de aminoácidos presente nos seres vivos classificados de acordo com o grupo R.	23
Figura 5 – Ligação peptídica entre dois aminoácidos resultando na liberação de uma molécula de água.	24
Figura 6 – Características de uma típica ligação peptídica, com os valores considerados consenso para os ângulos e comprimentos de ligação.	25
Figura 7 – Estrutura primaria de uma proteína.	26
Figura 8 – Estrutura secundária no formato hélice- α	27
Figura 9 – Estrutura secundária no formato folha- β	27
Figura 10 – Estrutura terciaria da proteína PDB: 1MNH.	28
Figura 11 – Estrutura quaternária da Hemoglobina.	29
Figura 12 – Energia de interação entre dois átomos pela sua distância.	31
Figura 13 – Representação esquemática da ligação dipolo instantâneo–dipolo induzido.	32
Figura 14 – Desnaturação de proteínas utilizando a variação térmica e a concentração de hidrocloreto de guanidina [GdnHCl].	35
Figura 15 – Funil de enovelamento de uma proteína em função dos contatos. nativos “Q”	37
Figura 16 – Superfície de energia de uma proteína, em formato de funil, apresentado as rotas que a proteína segue até o seu estado nativo.	39

Figura 17 – Perfil de energia livre de uma proteína devido a variação térmica.	41
Figura 18 – Calor específico temperatura determinando o enovelamento e desenovelamento de uma proteína.	43
Figura 19 – Estrutura 3D das Proteínas Cold Shock na forma de fechadura grega.	45
Figura 20 – Alinhamento dos aminoácidos das proteínas homologas (Psicrofílica, Mesofílica, Termofílica e Hipertermofílica).....	46
Figura 21 – Contatos nativos (Q) em função do tempo. computacional.	52
Figura 22 – Fração nativa em função das médias dos contatos nativos (<Q>) e Capacidade Térmica pela temperatura em unidades reduzidas.....	53
Figura 23 – Energia livre em função dos contatos nativos (Q) na temperatura de enovelamento Tf.	55

LISTA DE TABELAS

Tabela 1 – Lista com a massa molecular dos 20 aminoácidos e código de 1 letra ou de 20	
Tabela 2 – Nomenclatura de acordo com a IUPAC e fórmula estrutural linear dos 20 aminoácidos naturais presentes nos seres vivos.....	21
Tabela 3 – Determinação do RMSD das proteínas Cold Shock Homologas.	47
Tabela 4 – Comparativo da identidade das proteínas Cold Shock em porcentagem.	48
Tabela 5 – Temperatura de enovelamento teórico e experimental.....	54

LISTA DE ABREVIATURAS E SIGLAS

3D	Três dimensões
Bs-CspB	Proteína <i>Cold Shock</i> da bactéria termofílica <i>Bacillus subtilis</i>
Bc-Csp	Proteína <i>Cold Shock</i> da bactéria mesofílica <i>Bacillus caldolyticus</i>
Lm-Csp	Proteína <i>Cold Shock</i> da bactéria psicofílica <i>Listeria monocytogenes</i>
Tm-Csp	Proteína <i>Cold Shock</i> da bactéria hipertermofílica <i>Thermotoga maritima</i>
PDB	Protein Data Bank
GROMACS	GRONingen MACHine for Chemical Simulations
RMSD	Raiz media do desvio Quadrático
RMN	Ressonância Magnética Nuclear
SMOG	Structure-Based Models in GROMACS
RG	Raio de Giro
VMD	Visual Molecular Dynamics
WHAM	Weighted Histogram Analysis Method

Lista de símbolos

A	Variável dinâmica
C_α	Carbono alfa
C_p	Calor específico à pressão constante
D	Estado desenovelado
E	Energia
F	Energia livre de Helmholtz
k_B	Constante de Boltzmann
ΔG	Energia livre de Gibbs
m	Massa
N	Estado nativo
n	Número de partículas do sistema
Q	Função de partição
q	Coordenadas generalizadas
S	Entropia
T	Temperatura
T_m	Temperatura de enovelamento
T_f	Temperatura de enovelamento
T_G	Temperatura de vidro
T_s	Temperatura de estabilidade máxima
β	Temperatura recíproca
ω	Quantidade de estados possíveis para um sistema.

SUMÁRIO

1. Objetivo	14
2. Revisão Bibliográfica	17
2.1. Proteínas e sua Composição	17
2.1.1. Aminoácidos	18
2.1.1.1. <i>Classificação dos Aminoácidos</i>	22
2.1.2. <i>Ligações Peptídicas</i>	24
2.1.3. <i>Classificação Estrutural das Proteínas</i>	25
2.2. Estabilização de uma proteína no estado nativo.....	29
2.2.1 Interações Hidrofóbico.....	30
2.2.2 Ligações de Hidrogênio.....	30
2.2.3. Ligações Dissulfeto.....	30
2.2.4. Pontes Salinas.....	31
2.2.5. Forças de Van der Waals.....	31
2.2.6. Forças de Dispersão de London.....	32
2.3. <i>Desnaturação e renaturação de uma proteína</i>	34
2.4. <i>O Mecanismo de Enovelamento de Proteínas</i>	35
2.5. <i>O estudo da termoestabilidade em proteínas: um panorama geral</i> ..	39
2.5.1. <i>Estabilidade das Proteínas Homólogas</i>	40
2.6. Curvas de estabilidade.....	40
2.7. O calor específico e a entalpia.....	42
2.8. <i>As proteínas Cold Shock</i>	44
3. <i>Materiais e Métodos</i>	48
3.1. Modelo Carbono- α	48
3.2. Simulação Computacional.....	49
3.2.1. Método Múltiplos Histogramas.....	50
4. Resultados e Discussão.....	52
5. Conclusão.....	56
6. Referências.....	57

1. Objetivo.

Proteínas são polímeros formados por sequências de 20 diferentes resíduos (19 de aminoácidos e um de iminoácido) que, ao interagir físico-quimicamente, formam uma estrutura tridimensional (3D) única (LESK; STERNBERG, 2008). O melhoramento de novas técnicas computacionais relacionadas as sequências de DNA, proporcionaram uma melhor compreensão das sequências proteicas, também conhecidas como estruturas primárias de proteínas. Estas sequências, traduzidas do DNA, podem ser obtidas do *GenBank* (BENSON *et al.*, 2012). As estruturas 3D de proteínas, por sua vez, podem ser obtidas do *Protein Data Bank* (PDB) (BERMAN *et al.*, 2000).

Atualmente, há cerca de 16 milhões de sequências de proteínas únicas, que não se repetem. Entretanto, no PDB, encontramos aproximadamente 140 mil estruturas 3D de proteínas. Ao eliminarmos a redundância, filtrando estruturas muito similares, obtemos aproximadamente 1.393 formas de dobramento espaciais diferentes (*PDB Statistics*, acessado em 26 de agosto de 2018).

As proteínas são grandes moléculas, formadas por centenas de ligações simples formando assim o seu esqueleto covalente. Como é possível a livre rotação entre várias dessas ligações, a proteína consegue, em princípio, assumir um número de conformações praticamente incontáveis. Entretanto, cada proteína tem uma função biológica e uma estrutura específica, sugerindo que cada uma delas tenha uma estrutura tridimensional única (NELSON; COX, 2014).

As proteínas estão em todos os processos que ocorrem em uma célula, exibindo assim uma grande diversidade de funções vitais ao organismo. A estrutura das proteínas é determinada por sua sequência de aminoácidos, variando apenas um aminoácido da cadeia principal, pode-se ter a total mudança de função da proteína.

A função da proteína depende exclusivamente de sua estrutura. Apresentam como principal força de estabilização as ligações peptídicas, que são ligações químicas que se estabelecem entre um grupo carboxila de um aminoácido e um grupo amina de outro aminoácido subsequente, com a liberação de uma molécula de água, ou seja, formadas a partir da desidratação ou também quebradas por hidrólise. Reação de síntese por desidratação que ocorre nas moléculas de aminoácidos, apesar de possuírem diversas variações apresentam um padrão de estruturas comuns, que ajudam a organizar o entendimento sobre a arquitetura das proteínas.

Apresentam estruturas proteicas que não são estáticas. Todas as proteínas passam por mudanças na conformação variando desde sutis até bastante significativas. Partes de muitas proteínas possuem estruturas não discerníveis. A ausência ou falta dessas para algumas proteínas é determinante para sua função biológica.

As proteínas homologas desempenham a mesma função em tecidos ou em espécies diferentes. Estas proteínas apresentam ligeiras diferenças estruturais na sua cadeia de aminoácidos, esses segmentos com sequência diferentes de aminoácidos em proteínas homólogas são chamados "segmentos variáveis", e geralmente não participam diretamente da atividade da proteína. Os segmentos idênticos das proteínas homólogas são chamados "segmentos fixos", e são fundamentais para o funcionamento bioquímico da proteína.

No contexto da estrutura de proteínas, o termo estabilidade pode ser definido como a tendência em manter a conformação nativa. Proteínas nativas são apenas marginalmente estáveis: o ΔG que separa os estados dobrados e não dobrados em proteínas comuns, sob condições fisiológicas, está na faixa de apenas 20 a 65 kJ/mol.

Uma dada cadeia polipeptídica pode, teoricamente, assumir inúmeras conformações e, como resultado, o estado não dobrado de uma proteína é caracterizado por um alto grau de entropia conformacional. Essa entropia, junto com as interações de ligações de hidrogênio dos diversos grupos da cadeia polipeptídica com o solvente (água), tende a manter o estado não dobrado. As interações químicas que contrabalançam esses efeitos e estabilizam a conformação nativa incluem ligações dissulfeto (covalentes) e interações fracas (não covalentes).

As proteínas são os instrumentos moleculares por onde as informações da genética são expressas. As proteínas de cada organismo, da mais simples das bactérias aos seres humanos, são construídas a partir do mesmo conjunto onipresente de 20 aminoácidos (NELSON; COX, 2014).

Pode-se sugerir a montagem aleatória de uma proteína através de uma determinada sequência de aminoácidos, mais não necessariamente, seja uma proteína. O resultado dessa sequência de aminoácidos é denominado como um heteropolímero aleatório. O heteropolímero aleatório é incapaz de satisfazer a maioria das suas interações energéticas simultaneamente e está incapacidade recebe o nome de frustração. Desta forma, o heteropolímero é classificado como sendo frustrado, não possuindo uma estrutura tridimensional estável e não realizando qualquer função biológica.

Os aminoácidos diferem umas das outras em suas cadeias laterais ou grupos R, que variam em estrutura, tamanho e carga elétrica, e que influenciam a solubilidade dos aminoácidos em água. (NELSON; COX, 2014).

A composição de aminoácidos das proteínas também é muito variável. Os 20 aminoácidos comuns quase nunca ocorrem em quantidades iguais em uma proteína. Alguns aminoácidos podem ocorrer apenas uma vez ou estar ausentes em determinado tipo de proteína; outros podem ocorrer em grande número. (NELSON; COX, 2014).

Para todas as proteínas de todos os organismos, as interações covalentes são especialmente importantes para o enovelamento das cadeias polipeptídicas em suas estruturas secundárias e terciárias. A associação de múltiplos polipeptídeos para formar estruturas quaternárias também tem como base estas interações fracas. (NELSON; COX, 2014).

Uma proteína pode assumir diversas configurações tridimensionais, porém está apenas desempenha suas funções quando está arranjada em um estado conformacional denominado nativo. Esse estado é estabilizado predominantemente pelas interações fracas, como as ligações de hidrogênio, as interações hidrofóbicas, iônicas, eletrostáticas, dentre outras. Porém, alguns tipos dessas interações também podem assumir um caráter desestabilizador dependendo da situação tratada (CONTESSOTO *et al.*, 2017; NELSON; COX, 2014).

Esse processo de auto-organização que descreve as interações entre os constituintes das proteínas e com o meio onde ela está solvatada, esse processo ocorre por um caminho preferencial chamado de funil de enovelamento, termodinamicamente esse funil é caracterizado por uma energia livre de Gibbs em função de uma coordenada de reação. Com auxílio de simulações computacionais é possível determinarmos os perfis de energia livre e os coeficientes de difusão do processo para diferentes temperaturas, encontrando assim a temperatura de enovelamento.

Este trabalho tem como objetivo compreender o estudo das proteínas por meio de simulações computacionais, bem como identificar termoestabilidade das proteínas Cold Shock homólogas (SANTOS FILHO; ALENCASTRO, 2003; SILVA, 2005; PERL *et al.*, 1998; PHADTARE; ALSINA; INOUE, 1999; ULUSU; TEZCAN, 2001).

Compreender os mecanismos que proporcionam a termoestabilidade em proteínas homólogas pode ser de grande valia para diversas aplicações. A partir disso, seria possível a realização de modificações em proteínas com a finalidade de adaptá-las para

desempenhar suas funções em condições desejadas que originalmente não seriam possíveis ou não muito eficientes. Nesse contexto, pode se encaixar, por exemplo, o desenvolvimento de enzimas aperfeiçoadas para o branqueamento de papel e a produção do bioetanol para tornar mais viável e eficiente seu uso (CONTESSOTO, 2016; STERPONE; MELCHIONNA, 2012).

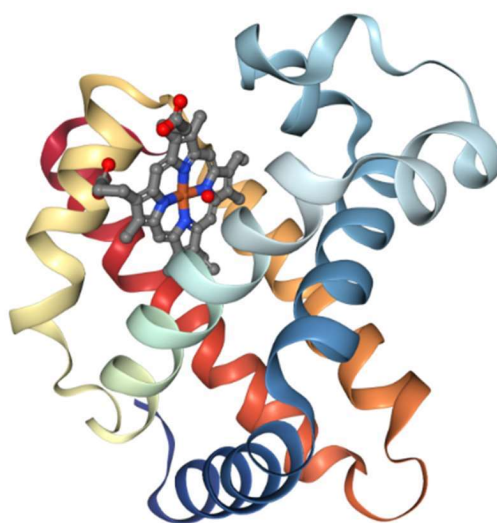
2. Revisão Bibliográfica

Este capítulo busca apresentar uma base teórica referente aos principais conceitos que vão nortear o trabalho. Abordando o conceito de proteínas, o problema das variações de temperaturas das proteínas *Cold Shock*, problema da variação dos resíduos de aminoácidos. A utilização dos principais métodos computacionais de simulação de proteínas, seguidos de conceitos relacionados à simulação termodinâmica e análise dos resultados obtidos dessas simulações.

2.1. Proteínas e sua Composição

Proteínas são cadeias formadas por polipeptídeos, que são polímeros lineares constituídos de resíduos de aminoácidos. Esses polímeros, ou biopolímeros, são macromoléculas caracterizadas pela repetição de unidades menores em sua formação, chamadas de monômeros. A Figura 1 mostra a Mioglobina, a primeira proteína a ter a sua estrutura resolvida utilizando-se a técnica de difração de raios X (KENDREW *et al.*, 1958). Em geral, é muito comum que os termos proteínas e polipeptídeos sejam usados indistintamente, assim como aminoácidos e resíduos de aminoácidos. Contudo, moléculas referidas como polipeptídeos possuem uma massa molecular abaixo de 10.000 u, enquanto que as proteínas possuem massas moleculares maiores. (CAMARGO, 2007).

Figura 1 – Estrutura quaternária da mioglobina, primeira proteína a ter sua estrutura determinada por raio-X.

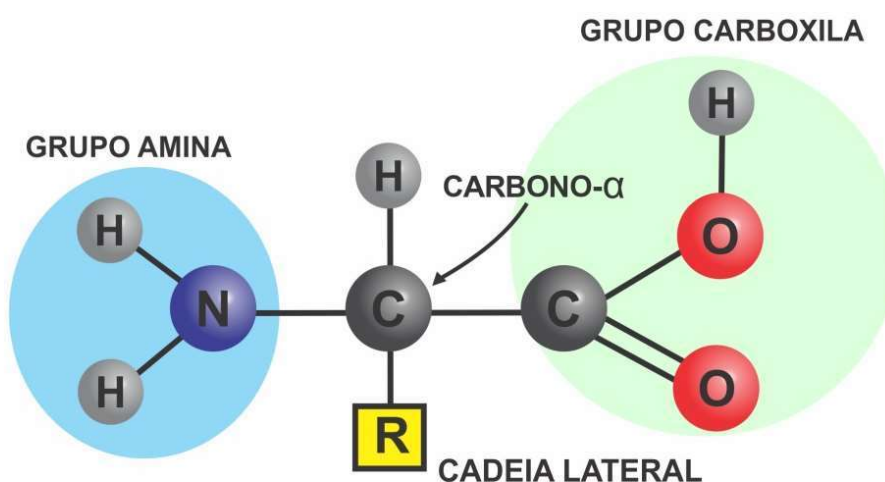


(Fonte: RCSB Protein Data Bank (2018))

2.1.1. Aminoácidos

Aminoácidos são moléculas orgânicas que contêm um grupo amina, um grupo carboxila e uma cadeia lateral R (específica para cada aminoácido) ligadas a um mesmo carbono, denominado carbono-alfa, ou carbono- α , ou C_{α} , a figura 2 apresenta uma estrutura típica de um aminoácido.

Figura 2 – Estrutura típica de um aminoácido, formado por um grupo amina, um grupo carboxila e um radical.

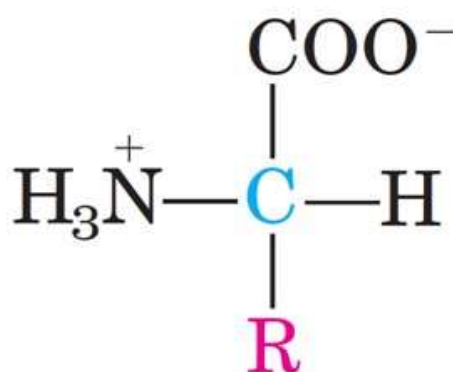


(Fonte: ALMEIDA, (2016))

A composição geral dos aminoácidos é feita de carbono, oxigênio, hidrogênio e nitrogênio. Os aminoácidos se ligam através de ligações covalentes chamadas de ligações peptídicas, liberando uma molécula de água durante esta reação, chamada de reação de condensação, que é uma classe comum de reações em células vivas.

Por esta razão, é utilizado o termo resíduos de aminoácidos para fazer referência aos aminoácidos que perderam esta molécula de água. Não obstante, frequentemente os termos aminoácidos e resíduos de aminoácidos são usados indistintamente. Em geral, os aminoácidos que formam proteínas são denominados de alfa-aminoácidos (ou α -aminoácidos) e possuem a seguinte fórmula geral:

Figura 3 – Estrutura de um alfa-aminoácidos na forma de íon dipolar.



(Fonte: BAYNES, (2015)).

O primeiro aminoácido descoberto foi a Asparagina em 1806 e, somente depois em 1938, foi descoberto o último, a Treonina. Os aminoácidos diferem uns dos outros pela sua cadeia lateral R, que pode variar em estrutura, tamanho e carga elétrica.

A Tabela 1 lista os 20 aminoácidos naturais conhecidos com os seus respectivos códigos de identificação. O código de uma letra foi idealizado por Margaret Oakley Dayhoff (1925 – 1983). Já na Tabela 2 apresenta a nomenclatura dos aminoácidos de acordo com a IUPAC e a estrutura química linear de cada um deles.

Tabela 1 – Lista com a massa molecular dos 20 aminoácidos e código de 1 letra ou de 3 letras.

Aminoácidos	Código de 3 letras	Código de 1 letra	Peso Molecular (g/mol)
Alanina	Ala	A	89.0935
Cisteína	Cys ou Cis	C	121.1590
Ácido Aspártico ou Aspartato	Asp	D	133.1032
Ácido Glutâmico ou Glutamato	Glu	E	117.1299
Fenilalanina	Phe ou Fe	F	165.1900
Glicina ou Glicocola	Gly, Gli	G	75.0669
Histidina	His	H	155.1552
Isoleucina	Ile	I	131.1736
Lisina	Lys ou Lis	K	116.1882
Leucina	Leu	L	131.1736
Metionina	Met	M	149.2124
Asparagina	Asn	N	132.1181
Prolina	Pro	P	115.1310
Glutamina	Gln	Q	146.1451
Arginina	Arg	R	171.2017
Serina	Ser	S	105.0930
Treonina	Thr ou The	T	119.1197
Valina	Vai	V	117.1169
Triptofano	Trp ou Tri	W	204.2262
Tirosina	Tyr ou Tir	Y	181.1891

(Fonte: (ALMEIDA, 2016))

Tabela 2 – Nomenclatura de acordo com a IUPAC e fórmula estrutural linear dos 20 aminoácidos naturais presentes nos seres vivos.

Aminoácidos	Nomenclatura IUPAC	Estrutura linear / Fórmula Química
Alanina	2-aminopropiônico ou 2-amino-propanóico	CH ₃ -CH(NH ₂)-COOH / C ₃ H ₇ NO ₂
Cisteína	2-bis-(2-amino-propionico)-3-dissulfeto ou 3-tiol-2-amino-propanóico	HS-CH ₂ -CH(NH ₂)-COOH / C ₃ H ₇ NO ₂ S
Ácido Aspártico ou Aspartato	2-aminossuccínico ou 2-amino-butanodióico	HOOC-CH ₂ -CH(NH ₂)-COOH / C ₄ H ₇ NO ₄
Ácido Glutâmico ou Glutamato	2-aminoglutárico	HOOC-(CH ₂) ₂ -CH(NH ₂)-COOH / C ₅ H ₉ NO ₄
Fenilalanina	2-amino-3-fenil-propionico ou 2-amino-3-fenil-propanóico	Ph-CH ₂ -CH(NH ₂)-COOH / C ₉ H ₁₁ NO ₂
Glicina	2-aminoacético	NH ₂ -CH ₂ -COOH / C ₂ H ₅ NO ₂
Histidina	2-amino-3-imidazolpropionico	CH(NH ₂)-COOH / C ₆ H ₉ N ₃ O ₂
Isoleucina	2-amino-3-metil-n-valérico ou 2-amino-3-metil-pentanóico	CH ₃ -CH ₂ -CH(CH ₃)-CH(NH ₂)-COOH / C ₆ H ₁₃ NO ₂
Lisina	2, 6-diaminoexanóico	H ₂ N-(CH ₂) ₄ -CH(NH ₂)-COOH / C ₆ H ₁₄ N ₂ O ₂
Leucina	2-amino-4-metil-pentanóico	(CH ₃) ₂ -CH-CH ₂ -CH(NH ₂)-COOH / C ₆ H ₁₃ NO ₂
Metionina	2-amino-3-metiltio-n-butírico	CH ₃ -S-(CH ₂) ₂ -CH(NH ₂)-COOH / C ₅ H ₁₁ NO ₂ S
Asparagina	2-aminossuccionâmico	H ₂ N-CO-CH ₂ -CH(NH ₂)-COOH / C ₄ H ₈ N ₂ O ₃
Prolina	pirrolidino-2-carboxílico	NH-(CH ₂) ₃ -CH-COOH / C ₅ H ₉ NO ₂
Glutamina	2-aminoglutarámico	H ₂ N-CO-(CH ₂) ₂ -CH(NH ₂)-COOH / C ₅ H ₁₀ N ₂ O ₃
Arginina	2-amino-4-guanidina-n-valérico	HN=C(NH ₂)-NH-(CH ₂) ₃ -CH(NH ₂)-COOH / C ₆ H ₁₄ N ₄ O ₂
Serina	2-amino-3-hidroxi-propanóico	HO-CH ₂ -CH(NH ₂)-COOH / C ₃ H ₇ NO ₃
Treonina	2-amino-3-hidroxi-n-butírico	CH ₃ -CH(OH)-CH(NH ₂)-COOH / C ₄ H ₉ NO ₃
Valina	2-amino-3-metil-butanóico	(CH ₃) ₂ -CH-CH(NH ₂)-COOH / C ₅ H ₁₁ NO ₂

(Fonte: (ALMEIDA, 2016)).

2.1.1.1. Classificação dos Aminoácidos.

Os aminoácidos estão classificados em cinco grupos, em função da estrutura de seus grupos residuais (grupos R) de acordo com a polaridade de carga (GONZALEZ; SILVA, 2017).

(a) Aminoácidos Apolares (Gly, Ala, Val, Leu, Ile, Pro): seus grupos R são alifáticos e hidrofóbicos; a glicina é o aminoácido mais simples e a prolina é um iminoácido (grupo amina secundário), pois o carbono α está unido com o extremo do grupo R, ciclizando a molécula e deixando-a mais rígida (GONZALEZ; SILVA, 2017).

(b) Aminoácidos polares sem carga (Ser, Cys, Met, Asn, Gin): são hidrofílicos e sua polaridade pode ser dada pelos grupos hidroxila, amida ou sulfidril (tiol), que formam pontes de H com a água; asparagina e glutamina são amidas dos ácidos aspártico e glutâmico, respectivamente; a cisteína pode sofrer oxidação em seu grupo sulfidril (SH) e formar um composto dimérico (Cys-Cys ou cistina) por união de duas cisteínas mediante uma ponte dissulfeto (S-S); essas pontes são comuns nas proteínas e contribuem para estabilizar a molécula. (GONZALEZ; SILVA, 2017).

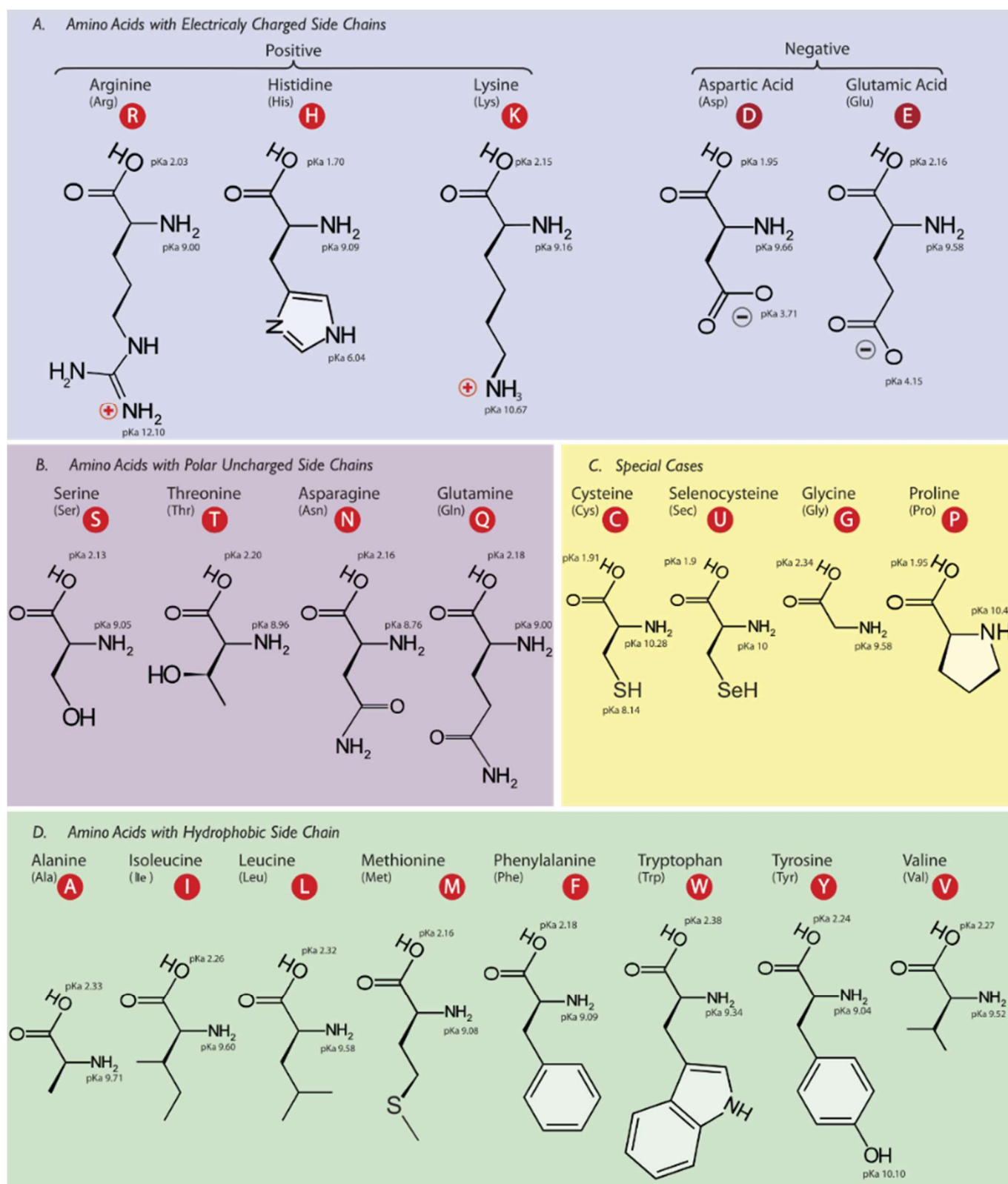
(c) Aminoácidos carregados negativamente ou aminoácidos ácidos (Asp, Glu): a carga está determinada pelos grupos carboxila ionizados. (GONZALEZ; SILVA, 2017).

(d) Aminoácidos carregados positivamente ou aminoácidos básicos (Lys, Arg, His): a carga positiva está determinada pelos grupos amina (Lys), guanidino (Arg) ou imidazol (His). (GONZALEZ; SILVA, 2017).

Todos os aminoácidos livres podem servir como tampões, assim como os aminoácidos que apresentam carga quando ligados às cadeias peptídicas. O carbono α de cada aminoácido (com exceção da glicina) está ligado a quatro diferentes grupos químicos e é, portanto, um átomo de carbono quiral ou opticamente ativo. Apenas a forma L dos aminoácidos é encontrada nas proteínas sintetizadas pelo corpo humano (HARVEY; FERRIER, 2015).

A figura 4 ilustra as estruturas químicas dos 20 aminoácidos, em suas fórmulas estruturais, ou seja, a disposição dos átomos que compõem os elementos químicos, bem como a ligação entre eles.

Figura 4 – Os 20 tipos de aminoácidos presente nos seres vivos classificados de acordo com o grupo R.

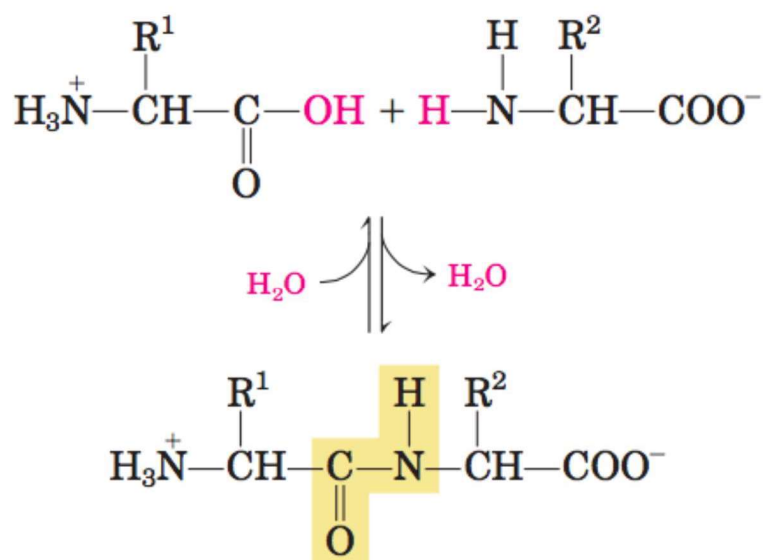


(FONTE: CORNELL et al., 1995)

2.1.2. Ligações Peptídicas

A ligação peptídica é um tipo de ligação covalente entre o grupo amina de um aminoácido com o grupo carboxila do outro. Quando esta ligação acontece, ocorre a liberação de uma molécula de água figura 5. Devido a essa perda, estes aminoácidos passam a ser chamados de resíduos de aminoácidos. À medida que muitos aminoácidos se unem, são formadas longas cadeias denominadas polipeptídeos e estes, por sua vez, formam as proteínas. Sabe-se que a composição das proteínas é constituída por milhares de polipeptídicos. (NELSON; LEHNINGER; COX, 2008).

Figura 5 – Ligação peptídica entre dois aminoácidos resultando na liberação de uma molécula de água.



(Fonte: NELSON; LEHNINGER; COX, (2008))

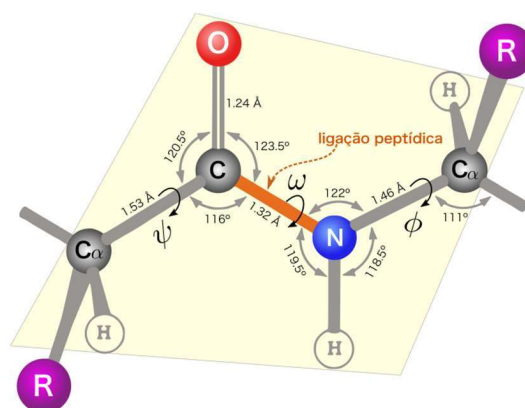
No final da década de 30, Linus Pauling e Robert Corey iniciaram estudos cristalográficos por difração de raios X da estrutura de aminoácidos e proteínas, cuja meta era a obtenção das distâncias e ângulos padrões das ligações entre os átomos, e a utilização desta informação na predição de conformações proteicas.

Em 1953, Miller conseguiu formar aminoácidos em laboratório (MILLER, 1953), substâncias formadoras de proteínas. Descobriu-se que a função de uma proteína depende da sequência de aminoácidos formados, pois uma mudança na estrutura de uma proteína pode levar a uma alteração da função. A natureza hierárquica da arquitetura de proteínas pode ser descrita em níveis de estruturas.

As ligações peptídicas C–N apresentam um caráter de ligação dupla, são rígidas e não podem rotacionar livremente, ao passo que não existem restrições para rotações entre os pares N–C α e C α –C, representadas pelos ângulos diedros Φ e ψ , respectivamente, que é ângulo formado pela intersecção de dois semiplanos com origem em uma mesma reta. Assim, a cadeia principal da ligação peptídica, também conhecida como backbone, está situada em uma série de planos rígidos onde todos os C α adjacentes são coplanares, compreende toda a cadeia de átomos da ligação peptídica excetuando-se as cadeias laterais.

A Figura 6 mostra os três ângulos diedros, Φ , ψ , e ω do backbone (também conhecidos como ângulos torcionais), além dos comprimentos das ligações e dos ângulos entre os átomos, onde são indicados os valores considerados padrões pela comunidade científica. O plano em amarelo ilustra a área que compreende todo o backbone. Os ângulos diedros formados com a cadeia lateral são chamados de χ (lê-se “chi”) e variam de χ_1 a χ_5 de acordo com cada ligação sucessiva ao longo da cadeia lateral (ALMEIDA, 2016).

Figura 6 – Características de uma típica ligação peptídica, com os valores considerados consenso para os ângulos e comprimentos de ligação.



(Fonte: ALMEIDA, (2016))

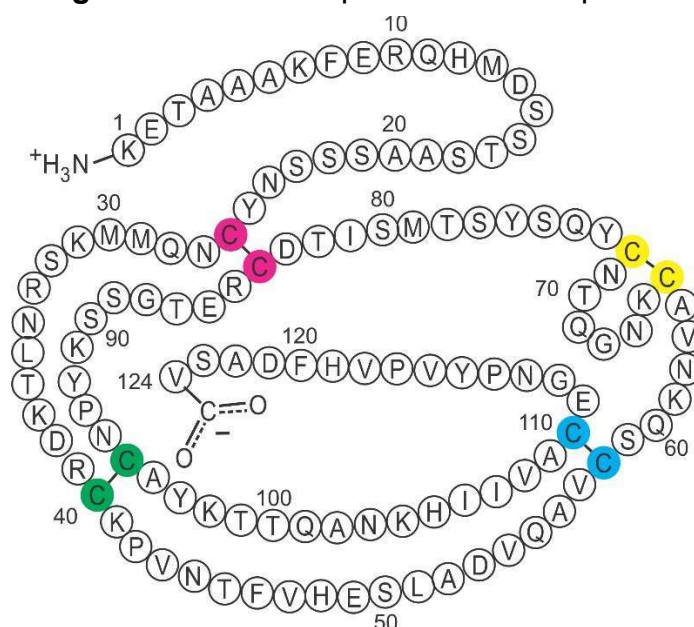
2.1.3. Classificação Estrutural das Proteínas

Proteínas podem ser agrupadas de acordo com suas similaridades sequenciais e/ou estruturais. Para a classificação sequencial, são utilizadas as técnicas de alinhamento entre sequências a fim de encontrar padrões de aminoácidos conservados. Na classificação

estrutural, os tamanhos e arranjos das estruturas secundárias de novas proteínas são comparados com estruturas tridimensionais conhecidas (DURHAM, 2007).

Estrutura Primária: A estrutura primária é dada pela sequência de aminoácidos ao longo da cadeia polipeptídica. É o nível estrutural mais simples e mais importante, pois dele deriva todo o arranjo espacial da molécula. São específicas para cada proteína, sendo geralmente determinados geneticamente. A estrutura primária da proteína resulta em uma longa cadeia de aminoácidos semelhante a um "colar de contas", com uma extremidade "amino terminal" e uma extremidade "carboxi terminal". Sua estrutura é somente a sequência dos aminoácidos, sem se preocupar com a orientação espacial da molécula.

Figura 7 – Estrutura primaria de uma proteína.

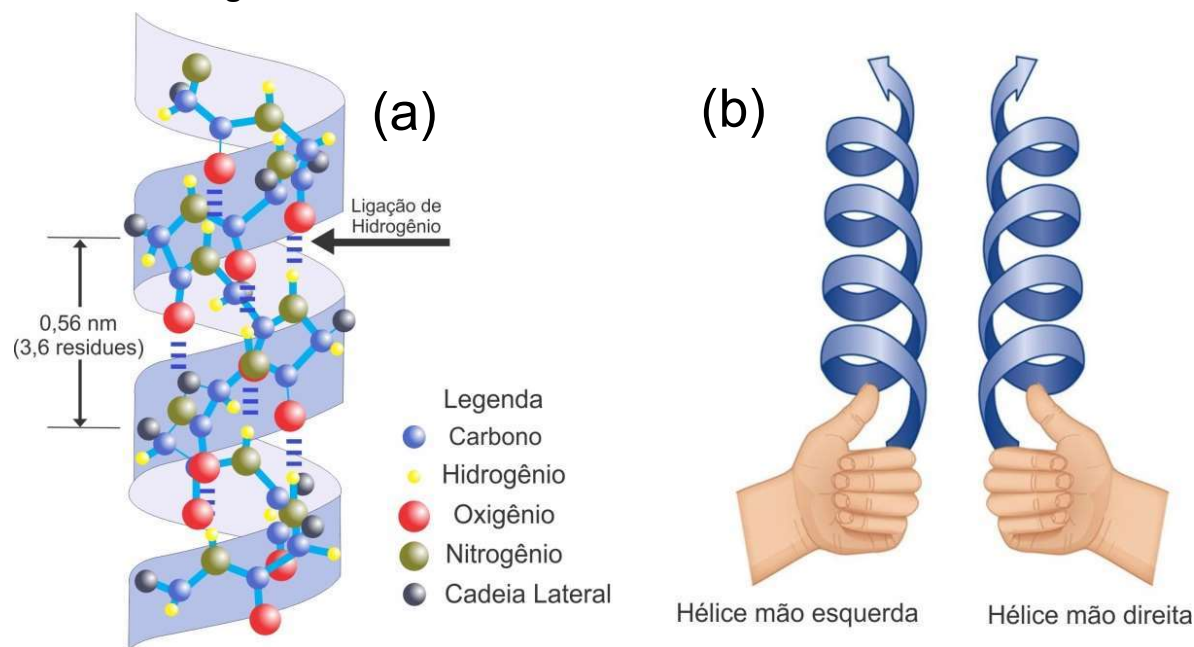


(Fonte: SANTOS FILHO, 2003)

Estrutura Secundária: A estrutura secundária refere-se à escolha de qualquer segmento de um polipeptídeo e descreve o arranjo espacial local da cadeia principal, sem considerar as conformações das cadeias laterais ou a sua relação com outros segmentos. Uma característica importante das estruturas secundárias é que os diedros e da cadeia principal repetem-se em padrões regulares, ou aproximadamente regulares, ao longo de todo o segmento considerado.

Existem dois tipos de estruturas secundárias muito comuns e estáveis que ocorrem em quase todas as proteínas: as hélices- α (figura 8) e as folhas- β (figura 9).

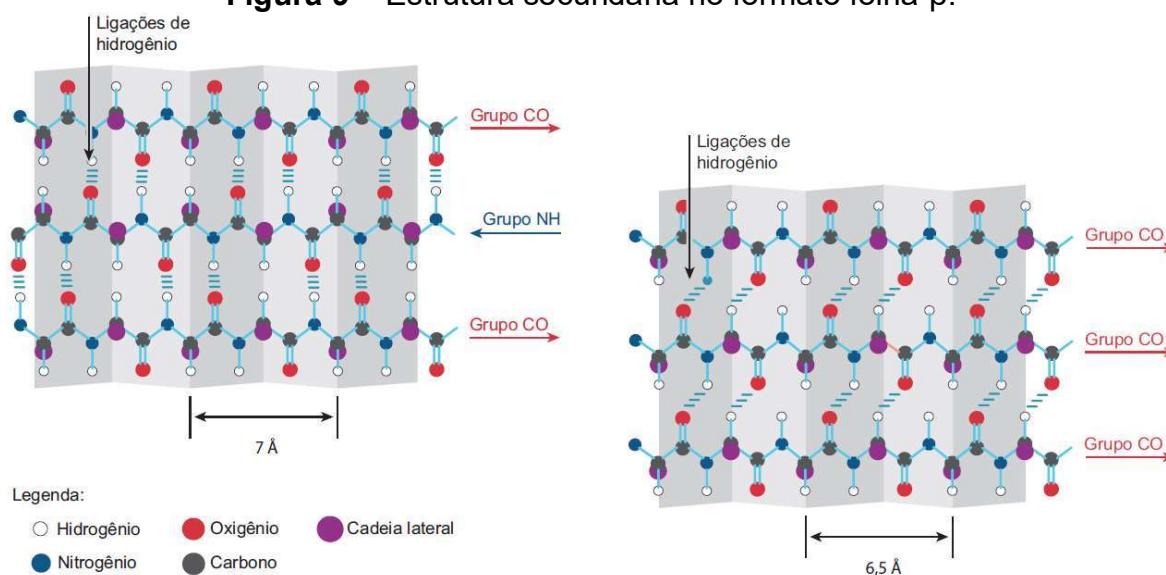
Figura 8 – Estrutura secundária no formato hélice- α .



(Fonte: ALMEIDA, 2016)

- (a) Modelo de bola-bastão com as indicações das ligações de hidrogênio.
 (b) Sentidos possíveis da orientação da volta da hélice: mão esquerda e mão direita.

Figura 9 – Estrutura secundária no formato folha- β .



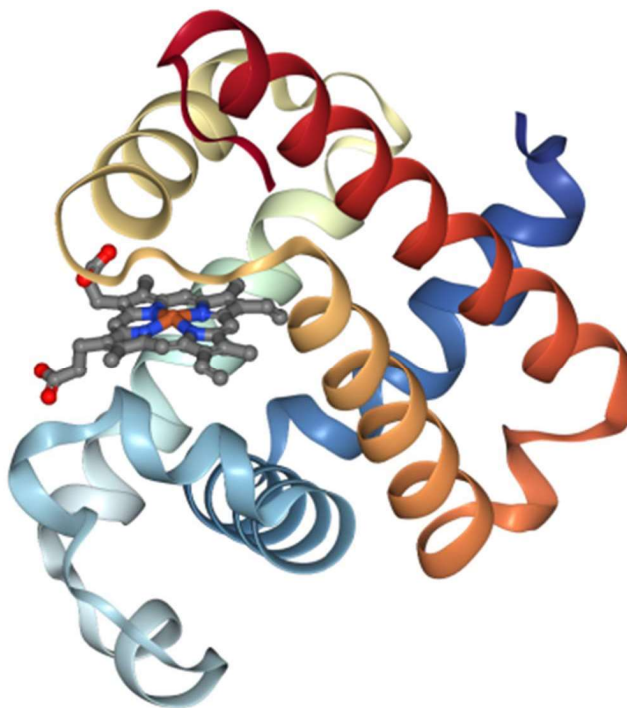
(a) Orientação antiparalela.

(b) Orientação paralela.

(Fonte: ALMEIDA, 2016)

Estrutura Terciária: A estrutura terciária de uma proteína corresponde à sua disposição tridimensional; ou seja, é o enrolamento dos elementos da estrutura secundária, juntamente com a disposição espacial das suas cadeias laterais. A primeira estrutura de proteína determinada por raios-X (mioglobina do esperma de baleia) foi elucidada no final dos anos de 1950 por John Kendrew (1917 – 1997) e pelos seus colaboradores. A cadeia polipeptídica desta proteína foi de tal difícil compreensão que estes investigadores reportaram o seu desapontamento e frustração sobre ao não conseguirem obter uma estrutura regular. Nos anos seguintes, documentaram-se mais de 18000 estruturas de proteínas, em que cada uma delas possui uma estrutura única e altamente complexa (CHOTHIA, 1990).

Figura 10 – Estrutura terciária da proteína PDB: 1MNH

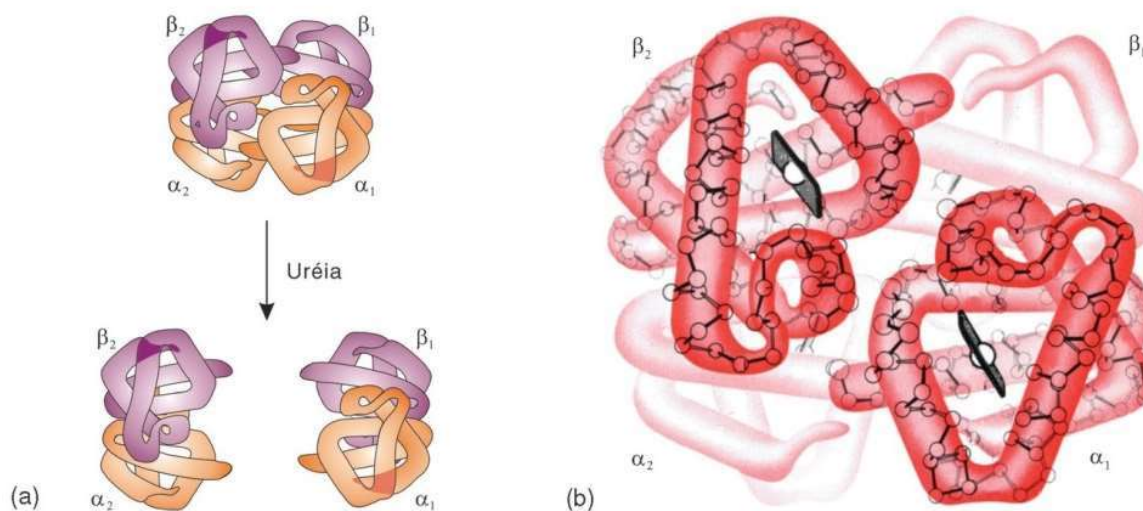


(Fonte: RCSB Protein Data Bank (2018)).

Estrutura Quaternária: Muitas proteínas consistem de duas ou mais cadeias polipeptídicas com estrutura terciária bem definida, consideradas subunidades da proteína. Esta organização de subunidades de cadeias polipeptídicas independentes constitui a estrutura quaternária de uma proteína (FISCHER, 2006). Algumas das vantagens de estruturas quaternárias são, maior estabilidade, eficiência genética (é necessário muito menos DNA para gerar um oligômero constituído por subunidades iguais do que diferentes de mesma massa), junção de sítios catalíticos (o monômero por si só pode não funcionar como um

sítio de ação enzimático completo, porém um oligômero teria esta capacidade) e cooperatividade (FISCHER, 2006).

Figura 11 – Estrutura quaternária da Hemoglobina.



(Fonte: MOREIRA, (2011))

Estrutura quaternária da hemoglobina. a) A associação entre subunidades diferentes é mais forte que entre subunidades iguais: quando a hemoglobina é suspensa em solução a concentração de ureia, o tetrâmero dissocia-se em dímeros $\alpha\beta$; os grupos heme não estão mostrados. b) Esquema enfocando as áreas de contato entre os dímeros, com a interface $\alpha_1\beta_1$ à frente; cada grupo heme fica alojado em um bolsão hidrofóbico. O reconhecimento biomolecular geralmente ocorre devido a complementaridade entre o ligante e a proteína.

2.2. Estabilização de uma proteína no estado nativo.

A estabilidade da proteína depende das principais forças indutoras do processo de enovelamento. As interações não-covalentes, tais como as ligações de hidrogênio, as ligações iônicas (pontes salinas), o efeito hidrofóbico e as interações de van der Waals (por exemplo, a força de dispersão de London), são bem mais fracas do que as ligações covalentes, entretanto, devido ao fato de que ocorrem inúmeras vezes, o efeito cumulativo garante a sua predominância para a estabilidade estrutural.

O termo estabilidade pode ser definido como a tendência de manter a conformação nativa (NELSON; COX, 2014). Contudo, as proteínas em sua estrutura nativa são fracamente estáveis, a diferença de energia livre de Gibbs (ΔG) separando um estado enovelado de um não-enovelado está entre 20 a 65 kJ/mol apenas. Para macromoléculas, a estrutura mais estável, ou seja, a estrutura nativa, é o estado em que a ocorrência de

interações fracas é máxima. A única interação covalente que influencia significativamente no processo de folding são as pontes dissulfetos.

2.2.1 Interações Hidrofóbico.

Durante o processo de folding, o efeito hidrofóbico é considerado o mais importante. O efeito consiste na tendência apresentada por substâncias apolares de se agregarem quando imersas em soluções aquosas, repelindo a água. Aminoácidos hidrofóbicos, ou que apresentam uma parte hidrofóbica (anfipáticos), agrupam-se no interior da proteína, formando um núcleo hidrofóbico no centro da proteína enovelada, enquanto a superfície externa contém a maioria dos resíduos polares (LI; TANG; WINGREEN, 1997) demonstraram, matematicamente, que o efeito hidrofóbico origina a principal força indutora no mecanismo de enovelamento.

2.2.2 Ligações de Hidrogênio.

As ligações de hidrogênio são predominantemente interações eletrostáticas (mas com ~10% de caráter covalente) entre um grupo dador (D-H) fracamente ácido e um aceitador (A) que possui um par de elétrons livres.

A ligação de hidrogênio acontece quando um grupo doador possui um átomo de hidrogênio que se liga a um átomo de alta eletronegatividade de um grupo receptor, como o flúor, nitrogênio ou oxigênio. Esse tipo de ligação pode ocorrer dentro da própria molécula (intramolecular), ou entre moléculas distintas (intermolecular). As ligações de hidrogênio favorecem a aproximação das cadeias laterais próximas aos grupos, aumentando o nível de empacotamento local.

2.2.3. Ligações Dissulfeto.

Uma vez que as ligações dissulfeto se formam quando a proteína se enrola para a sua conformação nativa, elas estabilizam a sua estrutura tridimensional. O caráter químico reduzido do citoplasma diminui a estabilidade das ligações intracelulares de dissulfeto.

Na verdade, praticamente todas as proteínas com ligações dissulfeto são secretadas para locais extracelulares que possuam ambientes mais oxidados, onde as suas ligações dissulfeto são mais eficazes em estabilizar a estrutura da proteína (ALBER, 1989).

2.2.4. Pontes Salinas.

As pontes salinas são interações iônicas que surgem entre uma cadeia lateral carregada positivamente e outra carregada negativamente. As pontes salinas de maior contribuição à estabilidade estrutural são aquelas formadas entre grupos de íons no núcleo hidrofóbico, favorecendo a criação de um ambiente essencialmente apolar. Isto auxilia fortemente na especificidade de uma conformação, desestabilizando aquelas nas quais as interações entre os íons não sejam ótimas (DILL; BROMBERG, 2012).

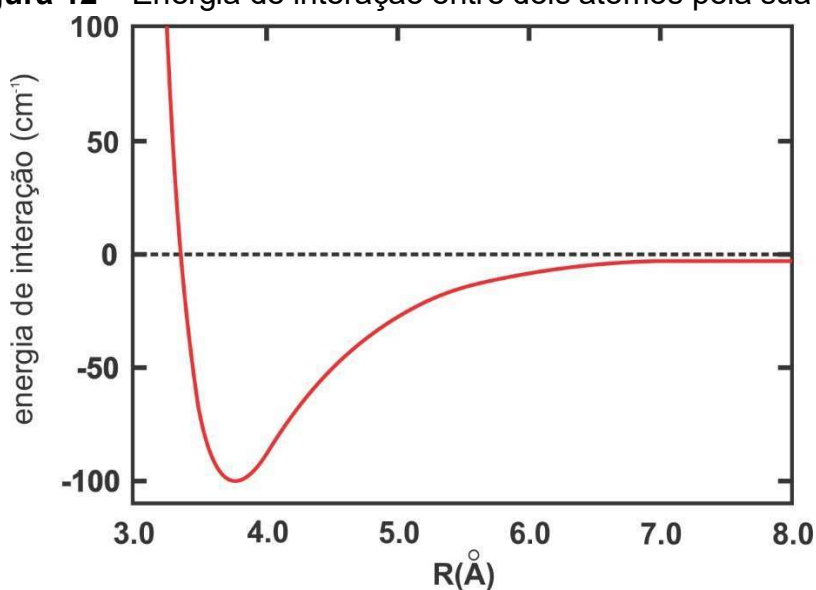
2.2.5. Forças de Van der Waals.

A força de Van der Waals consistem em interações intermoleculares, que não sejam devidas a ligações covalentes ou interações eletrostáticas entre íons. Caracterizam-se por apresentar o seguinte comportamento: para grandes distâncias a força apresenta um caráter atrativo, enquanto que para distâncias curtas a força tem caráter repulsivo. O potencial de Lennard-Jones, também conhecido como potencial L-J ou potencial 6-12, é frequentemente empregado para descrever este comportamento, cuja forma é:

$$U(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \quad (1)$$

Onde A e B representam um par de átomos neutros ou moléculas. A ligação de Van der Waals depende da distância entre os pares de átomos.

Figura 12 – Energia de interação entre dois átomos pela sua distância.



(Fonte: TEIXEIRA, (2001))

As forças dividem-se em três interações distintas:

1. Força entre dois dipolos permanentes: força de Keesom;
2. Força entre um dipolo permanente e um dipolo induzido: força de Debye;
3. Força entre dois dipolos instantaneamente induzidos: força de dispersão de London.

2.2.6. Forças de Dispersão de London.

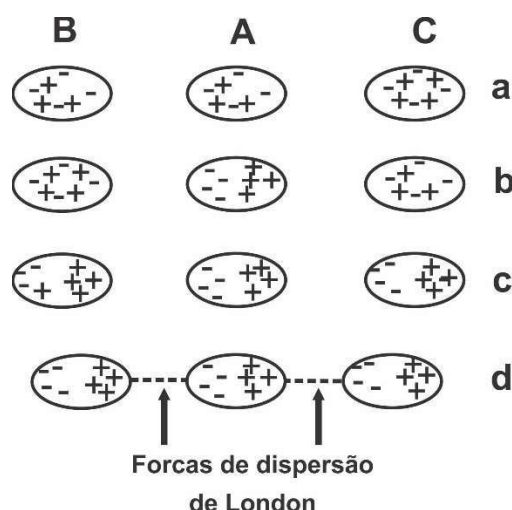
As ligações dipolo instantâneo - dipolo induzido ocorrem em moléculas apolares, quer monoatômicas (He, Ne, Ar, e.g.), quer poliatômicas (CCl₄, CH₄, N₂, H₂, e.g.).

Como os elétrons nas moléculas estão constantemente em movimento, num determinado instante podem acumular-se numa dada zona da molécula: forma-se um dipolo instantâneo, que pode, noutro instante, ter orientação diferente.

Esse dipolo, vai criar, em moléculas mais próximas, a formação de outros dipolos, designados dipolos induzidos: a zona do dipolo instantâneo com excesso de elétrons vai repelir elétrons de moléculas vizinhas; a parte com deficiência de elétrons vai atrair elétrons de moléculas vizinhas (RUSSO; SILVER, 2001).

Este processo prossegue entre as moléculas envolvidas. Entre as zonas de carga elétrica contrária surgem forças responsáveis pela ligação entre as moléculas. Estas forças são geralmente designadas por forças de dispersão de London (SALGUEIRO, 2003).

Figura 13 – Representação esquemática da ligação dipolo instantâneo–dipolo induzido.



(Fonte: SALGUEIRO, 2003)

- a. Moléculas apolares (A, B e C).
- b. Polarização instantânea de A.
- c. O dipolo instantâneo em A vai induzir dipolos em B e em C.
- d. Entre zonas de sinal contrário surgem interações responsáveis pela ligação intermolecular (—).

De um modo geral, a intensidade das forças de dispersão de London é proporcional ao número de elétrons e ao tamanho das moléculas e diminuem (no caso de isómeros) com o aumento de ramificação da cadeia. No caso de a ligação ocorrer entre moléculas iguais, a expressão da energia potencial é a seguinte (CHANG, 2000).

$$V = -\frac{3 \alpha^2 I}{4 r^6} \quad (2)$$

Onde α é a polarizabilidade, I a primeira energia de ionização da molécula e r a distância entre as moléculas.

O efeito da polarizabilidade de uma molécula na sua energia potencial de interação pode ser tal que, embora sendo, de um modo geral, este tipo de ligação intermolecular mais fraca que as restantes, pode acontecer que um composto apolar como, por exemplo, tetracloreto de carbono, CCl_4 , apresente um ponto de ebulição consideravelmente superior ao do flúor metano, CH_3F , polar, onde predominam ligações dipolo - dipolo: $76,5^\circ \text{ C}$ e $-141,8^\circ \text{ C}$, respectivamente (Chang, 2000). Sendo a ligação entre moléculas diferentes, A e B, a energia potencial é dada pela equação seguinte, equação de London (SALGUEIRO, 2003).

$$V = -\frac{3}{2} \frac{I_A I_B}{I_A + I_B} \frac{\alpha_A \alpha_B}{r^6} \quad (3)$$

- I_A e I_B são as primeiras energias de ionização de A e B, respectivamente.
- α_A e α_B as polarizabilidades de A e B, respectivamente.
- r é a distância entre A e B.

A equação anterior mostra a dependência da energia potencial relativamente a r^6 : a energia potencial sofre uma redução drástica quando a distância entre as moléculas aumenta (SALGUEIRO, 2003).

2.3. Desnaturação e renaturação de uma proteína.

O estado desnaturado não necessariamente corresponde ao desdobramento completo da proteína e à randomização da conformação. Na maioria das condições, as proteínas desnaturadas existem como um conjunto de estados parcialmente dobrados (NELSON; COX, 2014).

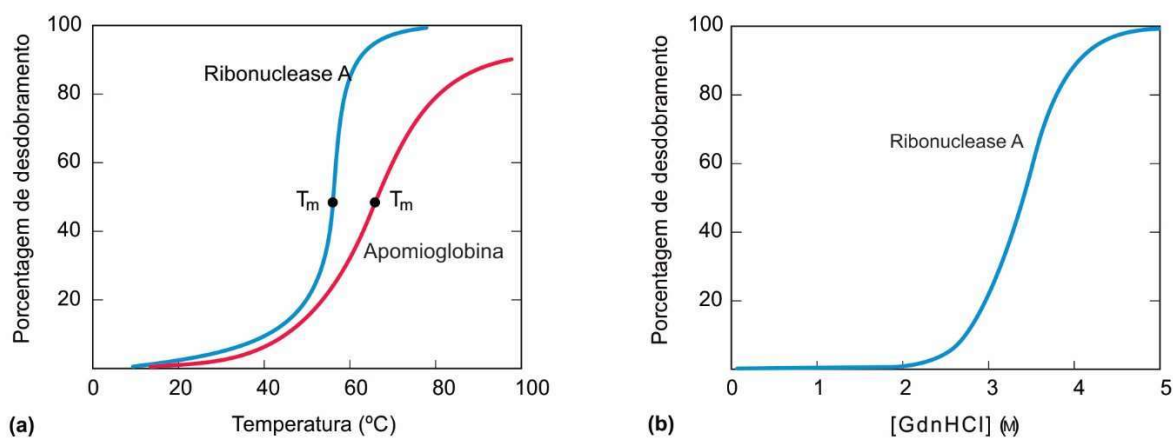
A maioria das proteínas pode ser desnaturada pelo calor, que tem efeitos complexos nas muitas interações fracas da proteína (principalmente sobre as ligações de hidrogênio) (NELSON; COX, 2014).

Se a temperatura é aumentada lentamente, a conformação da proteína em geral permanece intacta até que, em uma estreita faixa de temperatura, ocorre uma perda abrupta da estrutura e de função. A mudança repentina sugere que o desdobramento é um processo cooperativo: a perda de estrutura em uma parte da proteína desestabiliza as outras partes (NELSON; COX, 2014).

Os efeitos do aquecimento sobre as proteínas não são facilmente preditos. As proteínas altamente termorresistentes de bactérias termofílicas e de arqueobactérias evoluíram para funcionar na temperatura das fontes termais (100°C). Mesmo assim, as estruturas dessas proteínas diferem pouco de suas proteínas homólogas de bactérias como a *Escherichia coli*. Não se sabe ao certo como essas pequenas diferenças garantem a estabilidade estrutural em altas temperaturas (NELSON; COX, 2014).

As proteínas também podem ser desnaturadas por pH extremos, por certos solventes orgânicos miscíveis, como álcool ou acetona, por certos solutos como ureia e hidrocloreto de guanidina, ou por detergentes (Figura 14). Cada um desses agentes desnaturantes representa um tratamento relativamente brando, já que nenhuma ligação covalente da cadeia polipeptídica é rompida. Solventes orgânicos, ureia e detergentes atuam, principalmente, rompendo as interações hidrofóbicas que mantêm o núcleo estável das proteínas globulares; a ureia também rompe as ligações de hidrogênio; extremos de pH alteram a carga líquida da proteína, causando repulsão eletrostática e o rompimento de algumas ligações de hidrogênio. As estruturas desnaturadas resultantes desses vários tratamentos não são, necessariamente, as mesmas (NELSON; COX, 2014).

Figura 14 – Desnaturação de proteínas utilizando a variação térmica e a concentração de hidrocloreto de guanidina [GdnHCl].



(Fonte: ALMEIDA, 2016)

Desnaturação de proteínas. Os resultados correspondem a proteínas desnaturadas por duas modificações diferentes em seus ambientes. Em cada caso, a transição entre o estado dobrado e o estado não dobrado é abrupta, sugerindo cooperatividade neste processo. **(a)** Desnaturação térmica da apomioglobina (mioglobina sem o grupo prostético heme) de cavalo e da ribonuclease A (com suas ligações dissulfeto intactas). O ponto médio da faixa de temperatura em que ocorre a desnaturação é chamado de temperatura de fusão, ou T_m . A desnaturação da apomioglobina foi monitorada por dicroísmo circular, que mede a quantidade de estrutura helicoidal na proteína. A desnaturação da ribonuclease A foi acompanhada pelo monitoramento das mudanças na fluorescência intrínseca da proteína, que é afetada pelas mudanças no ambiente dos resíduos de Trp. **(b)** Desnaturação da ribonuclease A, com suas ligações dissulfeto intactas, pelo hidrocloreto de guanidina (GdnHCl), monitorado por dicroísmo circular.

2.4. O Mecanismo de Enovelamento de Proteínas.

As proteínas são não só as macromoléculas biológicas mais abundantes, mas também as mais complexas em termos conformacionais e biologicamente versáteis.

São sintetizadas como sequências lineares de aminoácidos a partir da informação contida no genoma nuclear, ou a partir do DNA mitocondrial no caso das 13 proteínas mitocondriais da cadeia respiratória. Após sintetizadas necessitam adquirir uma estrutura tridimensional específica, da qual depende diretamente a sua função, por um processo designado de folding proteico.

O processo de enovelamento proteico tem sido um assunto intensamente estudado desde os trabalhos pioneiros de Anfinsen e seus colaboradores acerca do folding/unfolding da ribonuclease.

Estes mostraram que uma cadeia polipeptídica pode adquirir o enovelamento espontaneamente *in vitro* sem requerer a assistência de nenhum fator celular, provando que toda a informação necessária para uma proteína adquirir a sua conformação nativa se encontra codificada na sua sequência linear de aminoácidos. Mostraram também que o

processo de folding das proteínas se baseia estritamente na termodinâmica, ou seja, na procura aleatória de uma conformação única termodinamicamente mais estável para a qual a energia livre seja mínima.

Todavia, Levinthal (1968) focando-se na cinética e dinâmica do folding, mostrou que o processo de folding não consiste na exploração de todas as conformações possíveis, mas sim ocorre através de uma via bem definida e sequencial. Caso contrário, até mesmo uma proteína pequena (~100 aminoácidos) necessitaria de um tempo biologicamente irrealista (~10²⁹ anos) para adquirir a sua estrutura nativa.

Na tentativa de se explicar o fato de as proteínas adquirirem a sua conformação nativa numa escala temporal de milissegundos, o folding proteico tem sido representado através de um funil de energia.

Este funil de energia tem sido usado para ilustrar que o grande número de conformações acessíveis para uma proteína (superfície do funil) se torna progressivamente mais restrito à medida que as interações nativas aumentam, terminando por fim numa estrutura nativa para a qual a energia livre é mínima (fundo do funil). Por oposição ao estado nativo, os estados unfolding são caracterizados por um alto grau de entropia conformacional e por uma alta energia livre de Gibbs (superfície do funil).

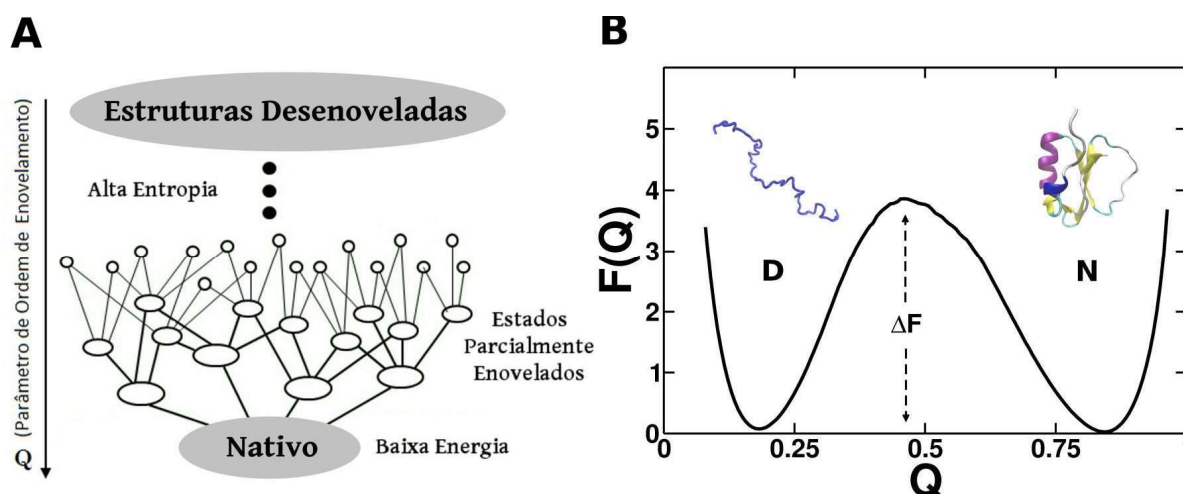
A teoria sobre a superfície de energia do enovelamento de proteína, introduzida por Onuchic, Wolynes e colaboradores (ZHURAVLEV; PAPOIAN, 2010; LEVINTHAL, 1968; POTOYAN; PAPOIAN, 2011; ITZHAKI; OTZEN; FERSHT, 1995), descreve e demonstra de forma simplificada o processo de enovelamento e desenovelamento de uma proteína e os principais mecanismos de transformação que ocorrem com as proteínas (CLEMENTI; NYMEYER; ONUCHIC, 2000; WALES, 2010; WALES, 2003; SHAN *et al.*, 2013; DOBSON, 2003). Ela se baseia na estatística do processo do enovelamento na qual são observados a superfície de energia global que uma proteína apresenta entre o seu desenovelamento e enovelamento (CHAVEZ; ONUCHIC; CLEMENTI, 2004; WANG *et al.*, 2012; REDDY *et al.*, 2010).

A superfície de energia tem caráter multidimensional e uma maneira simplificada de analisá-la é utilizando uma coordenada de reação. A coordenada de reação é necessária para acompanhar o processo de enovelamento ao longo da superfície de energia (WANG *et al.*, 2012) e a sua escolha é fundamental para uma boa abordagem ao problema estudado (WALES, 2010).

Uma coordenada de reação amplamente utilizada é a fração de contatos nativos Q . Esta coordenada é a razão entre os contatos nativos formados em uma conformação desenovelada D e os contatos nativos presentes na estrutura enovelada N . Desta forma, esta coordenada mede, ao longo do processo de enovelamento, o grau de similaridade entre uma estrutura desenovelada e o estado nativo. Os valores de Q variam entre 0 e 1, onde 0 significa que a proteína está totalmente desenovelada e 1 significa que a proteína está no seu estado nativo. A coordenada de reação Q é baseada na topologia do estado nativo da proteína e descreve de maneira satisfatória o processo de enovelamento (WANG *et al.*, 2012).

Nesta abordagem, o mecanismo de enovelamento pode ser entendido como um processo de difusão ao longo da superfície de energia quando se acompanha a coordenada de reação (DOYE; MILLER; WALES, 1999 - a, b).

Figura 15 – Funil de enovelamento de uma proteína em função dos contatos.



(Fonte: CONTESSOTO, 2018)

(a) Ilustração do funil de estruturas do processo de enovelamento de uma proteína ao longo da coordenada de reação Q . (b) Representação do perfil de energia livre $F(Q)$ em função da coordenada de reação Q . É destacado a existência do estado desenovelado D , do estado nativo N e da barreira de energia ΔF existente entre esses dois estados. Esta curva foi obtida por meio da simulação computacional do modelo $C\alpha$, para a proteína Cl2 - Chymotrypsin Inhibitor 2 PDB ID: 1YPA [4].

Na **Figura 15** estão descritas as estruturas da energia de superfície, quando vista sob a perspectiva de uma coordenada de reação dos contatos nativos, apresenta uma topologia afunilada com um gradiente de energia direcionado para a região do estado enovelado, ou seja, de menor energia. O conjunto de estruturas similares são agrupadas

ao longo do processo de enovelamento, onde a proteína passa por estados parcialmente enovelados e quantidade de estruturas possíveis diminui (ANFINSEN, 1973; WALES, 2012; ITZHAKI, OTZEN; FERSHT, 1995; XIAO; HONIG, 1999). Este processo ocorre até que a proteína atinja seu mínimo de energia na parte de baixo do funil de estruturas, atingindo o seu estado nativo. Pela descrição do funil de estruturas, não há um caminho preferencial no processo de enovelamento, mas uma multiplicidade de possíveis rotas (LEVINTHAL, 1968).

Desta perspectiva, durante o processo de enovelamento, existe uma competição entre a contribuição energética $E(Q)$ e a contribuição entrópica $S(Q)$, que são dependentes da coordenada de reação Q . O processo pode ser mapeado por um potencial efetivo que leve em consideração a competição entre essas duas grandezas. O potencial termodinâmico dado pela energia livre de Helmholtz ($F(Q)$) é utilizado para descrever esta competição, e é dado pela equação:

$$F(Q) = E(Q) - TS(Q). \quad (4)$$

com T sendo a temperatura do sistema.

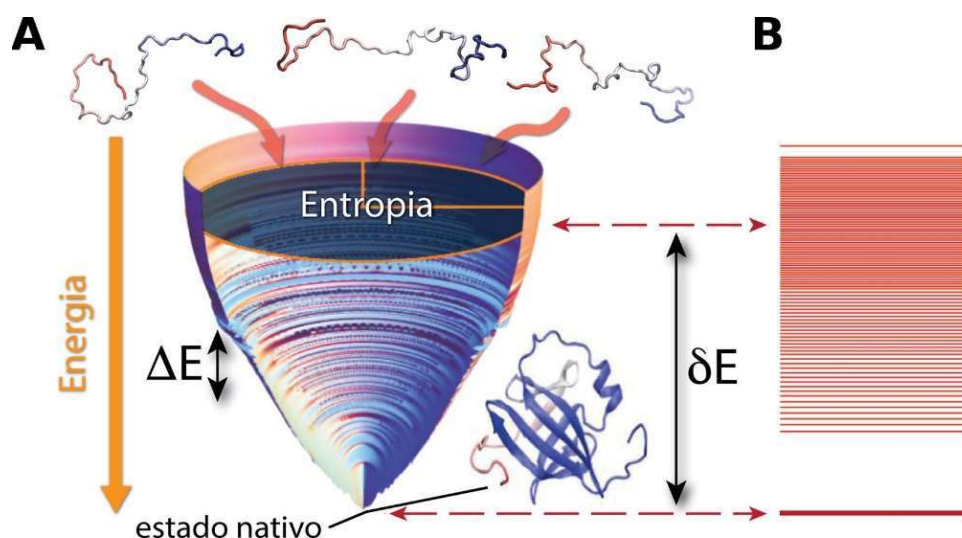
A figura 15B apresenta o perfil de energia livre ($F(Q)$) é construído em função da coordenada de reação Q para uma proteína típica com mecanismo de enovelamento de dois estados. A curva de energia livre destaca três regiões principais: o estado desenovelado D , o estado nativo N e a barreira de energia ΔF que existe entre os dois estados. A temperatura T utilizada na construção da curva é a temperatura em que ocorre enovelamento. Nesta temperatura, conhecida como temperatura de enovelamento T_f , a probabilidade da proteína se encontrar no estado desenovelado é a mesma de encontrá-la no estado nativo. Experimentalmente, a temperatura de enovelamento T_f é conhecida como a temperatura de enovelamento. Esta temperatura pode ser obtida por experimentos de calorimetria na construção de uma curva de desnaturação térmica ou de calor específico (DOYE; MILLER; WALES, 1999 - b; EVANS; WALES, 2003; KRIVOV; KARPLUS, 2002).

Estes modelos se distinguem, basicamente, pela identificação a priori de um determinado mecanismo. Evidentemente, os mesmos foram propostos baseados em observações experimentais que, devido à peculiaridade de cada experimento, enfatizam-se um ou outro mecanismo. Isto mostra, mais uma vez, a complexidade do problema do enovelamento proteico (SILVA, 2005).

As rugosidades da superfície de energia (Figura 16) também proporcionam transições de fase durante as mudanças de estados da proteína, dentre elas a transição de vidro. Nela, se a proteína não possui energia suficiente, possivelmente devido a uma condição de uma relativa baixa temperatura, ela fica aprisionada no respectivo estado por um maior período de tempo pela dificuldade em vencer a barreira de energia (OLIVEIRA, 2007; ONUCHIC; LUTHEY-SCHULTEN; WOLYNES, 1997).

A temperatura de vidro (TG) e a temperatura que delimita o aprisionamento da proteína em uma fase de vidro. Quanto mais distante dessa temperatura for a temperatura de enovelamento (T_f), temperatura em que a proteína possui 50% de chance de ser encontrada no estado nativo, mais chance a proteína possui de atingir o estado enovelado mais rapidamente. Porém, no caso da situação inversa, mais vagaroso seria esse processo (OLIVEIRA, 2007; ONUCHIC; LUTHEY-SCHULTEN; WOLYNES, 1997).

Figura 16 – Superfície de energia de uma proteína, em formato de funil, apresentado as rotas que a proteína segue até o seu estado nativo.



(Fonte: Adaptado de (WANG; OLIVEIRA *et al.*, 2012), p. 15765).

2.5. O estudo da termoestabilidade em proteínas: um panorama geral.

A termoestabilidade em proteínas ainda é uma questão em aberto. Diversos fatores que a influenciam foram identificados e sua interpretação pode ser dada em algumas perspectivas, porém a sua compreensão em plenitude ainda não foi atingida (STERPONE; MELCHIONNA, 2012).

Em uma perspectiva da análise termodinâmica, pode-se relacionar a estrutura e a composição das proteínas de modo a compreender melhor as interações que interferem na sua termoestabilidade. Proteínas mais estáveis possuiriam interações que tendessem a diminuir sua entalpia no estado nativo. Porém, em proteínas mais estáveis, como as termofílicas, a diferença de entropia em relação aos estados enovelado e desenovelado é baixa. Logo, compreender como a relação entalpia e entropia contribui para a termoestabilidade ainda é uma questão em aberto (STERPONE; MELCHIONNA, 2012).

Já em uma perspectiva cinética, pode-se associar a termoestabilidade das proteínas com a barreira de energia existente entre seu estado nativo e o desenovelado. Assim, quanto maior for essa barreira mais estável é a proteína. (STERPONE; MELCHIONNA, 2012).

2.5.1. Estabilidade das Proteínas Homólogas.

WELKER e colaboradores (1998) questionaram a diferença na estabilidade de proteínas Cold Shock homólogas provenientes de três organismos, da bactéria mesofílica *Bacillus subtilis* (Bs), Bs-CspB, da bactéria termofílica *Bacillus caldolyticus* (Bc), Bc-Csp, e da bactéria hipertermofílica *Thermotoga maritima* (Tm), Tm-Csp. Os autores apontam a grande similaridade entre as três proteínas em suas estruturas e cadeias de aminoácidos, e que, apesar disso, apenas a Tm-Csp e a Bc-Csp apresentam alta estabilidade, possuindo uma temperatura média de transição para desenovelamento igual a 85 e 72 °C, respectivamente, enquanto que a Bs-CspB é apenas marginalmente estável, com temperatura média para transição para o desenovelamento igual a 50°C.

Contudo a similaridade e identidade acentuada das proteínas Cold Shock, notamos que devido a alterações em apenas alguns aminoácidos da sua estrutura altera a estabilidade dessas proteínas, fazendo que seus estados Tf sejam alterados. Pode-se salientar também que devido a adaptação que essas bactérias podem sobreviver a condições físicas e químicas extremas, determinando assim a grande estabilidade dessas proteínas frente a variações de temperatura.

2.6. Curvas de estabilidade.

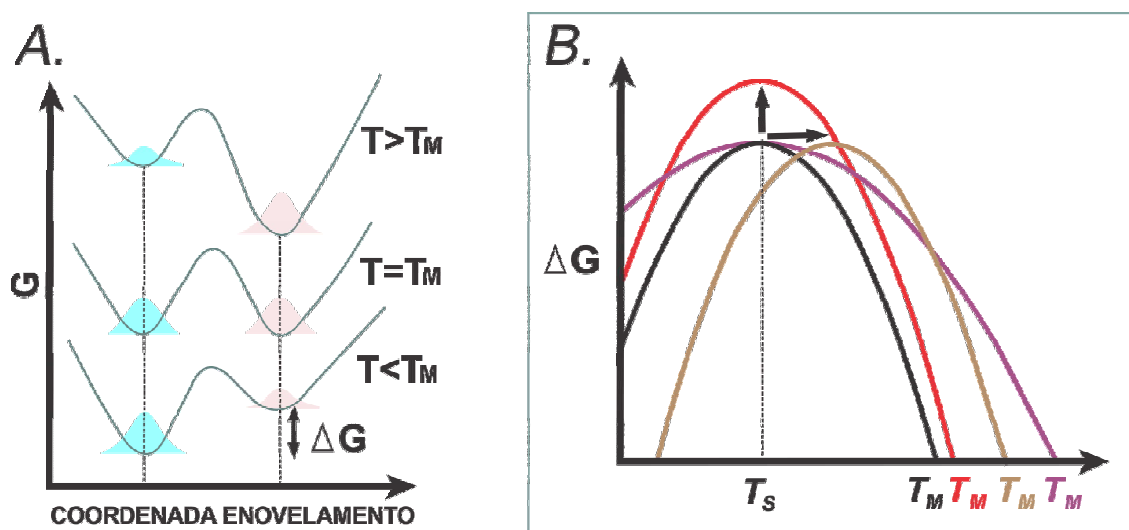
Experimentalmente, a estabilidade das proteínas pode ser analisada pelas curvas de estabilidade geradas em um processo de desenovelamento. Essas curvas são obtidas por

meio da variação da energia livre de Gibbs (G) no estado enovelado (N) e no estado desenovelado (D) (STERPONE; MELCHIONNA, 2012):

$$\Delta G = G(D) - G(N) \quad (5)$$

Sendo a denotação para $\langle D \rangle$ e $\langle N \rangle$ referindo-se à população em cada um dos estados (STERPONE; MELCHIONNA, 2012).

Figura 17 – Perfil de energia livre de uma proteína devido a variação



(Fonte: LUCAS, 2001)

A Figura 17 apresenta o perfil de energia livre de uma proteína em diferentes situações para T em relação a temperatura de enovelamento (T_m) durante o processo de desenovelamento, e curvas de estabilidade distintas para proteínas que possuem diferentes T_m . T_m possui interpretação análoga a T_F , representando a temperatura em que a população no estado nativo é equivalente a população no estado desenovelado. As curvas de estabilidade geralmente são mais largas nas proteínas com T_m mais elevado, resultado dessa maior estabilidade. A temperatura T_s representa a temperatura de máxima estabilidade de cada proteína (STERPONE; MELCHIONNA, 2012).

A função da curva de estabilidade (Figura 17) também pode ser representada por uma equação modificada que relaciona a energia de Gibbs e a energia de Helmholtz, é diversos outros parâmetros importantes no estudo da termoestabilidade (RAZVI; SCHOLTZ, 2006):

$$\Delta G(T) = \Delta H_m \left(1 - \frac{T}{T_m}\right) - \Delta C_p \left[(T_m - T) + T \ln \left(\frac{T}{T_m}\right) \right] \quad (6)$$

em que ΔH_m é a mudança de entalpia em T_m e ΔC_p é a variação do calor específico em relação ao processo de desenovelamento (RAZVI; SCHOLTZ, 2006).

Devemos ter em mente que a energia livre de Helmholtz é aplicada em processos a temperatura e volume constantes enquanto a de Gibbs é aplicada a temperatura e pressão constantes.

Pode ser salientado que quando desejamos trabalhar com apenas uma parte da energia do sistema utilizamos a energia de Helmholtz, pois mensura a parcela de energia interna de um sistema possível de ser utilizada na realização de trabalho, ou seja, é a energia útil que sobra para o sistema depois que utilizou parte da energia interna para expandir as fronteiras do sistema e redistribuir as moléculas nos diferentes níveis quânticos de energia, já se o interesse recair na energia total disponível para execução de trabalho útil, utilizamos a energia de Gibbs que engloba a energia total do sistema, onde temos a equação de Gibbs-Helmholtz para uma mudança de estado (transformação) que ocorre a pressão constante.

A integração da Equação de Gibbs-Helmholtz permite determinar a equação da energia livre de Gibbs (G) em função da temperatura (T) para o sistema mantido a pressão (P) constante. Para tanto, basta conhecer a função entalpia (H) em termos da temperatura. A mesma forma da equação que determina o valor da G para uma dada T, também se mantém para a determinação da variação de G (ΔG) em função da T para uma certa transformação, que ocorre a P constante.

$$d\left(\frac{\Delta G}{T}\right)_p = -\frac{\Delta H dT}{T^2} \quad (7)$$

2.7. O calor específico e a entalpia.

Experimentalmente, o calor específico (C_p) pode ser obtido por métodos de calorimetria de varredura diferencial em função da temperatura. Nessas técnicas, o processo pode ser estudado por meio da desnaturação e renaturação da proteína em estudo. A partir de dados do gráfico de C_p , como ΔC_p , é possível determinar a variação de entalpia de uma proteína no processo em questão (Figura 18). Além disso, também pode auxiliar na identificação do tipo de reação que ocorre no sistema em estudo, como, por exemplo, se esse é uma reação de dois estados (PRIVALOV; KHECHINASHVILI, 1974;

SABOURY; MOOSAVI-MOVAHEDI, 1994; STERPONE; MELCHIONNA, 2012). A função que descreve a entalpia é:

$$H = U + PV$$

$$dH = dU + pdV + vdP$$

$$dH = dQ - pdV + pdV + vdP$$

Para um processo isobárico (pressão constante) temos: (8)

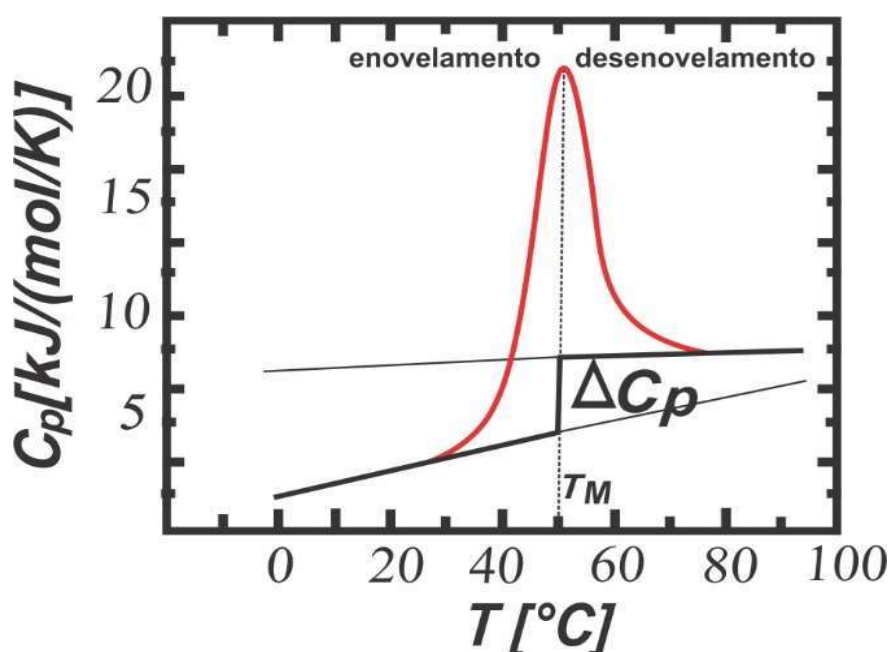
$$dH = dQ$$

$$\int dH = \int dQ$$

$$\Delta H = Q$$

Onde temos que quando $\Delta H > 0$, reação que absorve calor do meio (endotérmica), para $\Delta H < 0$ reação que libera calor para o meio externo (exotérmica).

Figura 18 – Calor específico temperatura determinando o enovelamento e desenovelamento de uma proteína.



(Fonte: Adaptado de (STERPONE; MELCHIONNA, 2012).

Uma das contribuições para a termoestabilidade proteica é a contribuição das interações eletrostáticas. Essa contribuição pode ser proveniente de interações de resíduos de aminoácidos carregados da proteína. As intensidades dessas interações são dependentes de diversos fatores, como pH em que a proteína é submetida, a área da proteína imersa no solvente, o nível de salinidade do solvente, a distribuição espacial desses aminoácidos carregados na estrutura da proteína, dentre outros. Evidenciando assim que a quantidade de resíduos carregados pode influenciar no enovelamento de uma proteína (CONTESSOTO *et al.*, 2016; CONTESSOTO *et al.*, 2017; NAKAMURA, 1996; STERPONE; MELCHIONNA, 2012; XIAO; HONIG, 1999). XIAO e HONIG (1999) ao analisar a contribuição das interações eletrostáticas na estabilidade de proteínas homólogas à pH constante observaram uma melhor contribuição nas proteínas hipertermofílicas frente as outras estudadas nesta dissertação, que geralmente são mais estáveis.

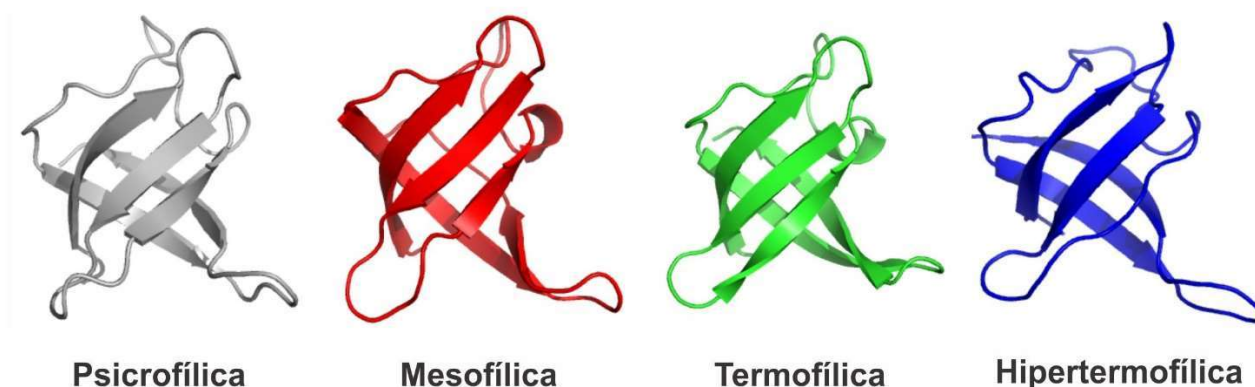
A partir de simulações computacionais, é possível realizar modificações nos grupos de aminoácidos carregados das proteínas buscando o aprimoramento dessa contribuição para, conseqüentemente, obter-se uma melhor termoestabilidade. O mesmo pode ser feito de maneira experimental, por meio de mutações nas proteínas (STERPONE; MELCHIONNA, 2012).

2.8. As proteínas Cold Shock.

As proteínas Cold Shock estudadas neste artigo presentes nas bactérias psicrófila *Listeria monocytogenes* (código PDB: 2LXJ) apresenta 66 resíduos e sua estrutura secundária é formada por 5 folhas *betas* e 4 laços, e atuam como chaperonas no RNA para regular a tradução (LEE *et al.*, 2013). Nas bactérias mesofílicas *Bacillus caldolyticus* (código PDB: 1CSP) apresenta 67 resíduos em sua estrutura secundária com 5 folhas *betas* e 4 laços, se ligam ao DNA por meio de ligações simples (SCHINDELIN; MARAHIEL; HEINEMANN, 1993). Nas termofílicas *Bacillus subtilis* (código PDB: 1C9O) apresenta 66 resíduos em estrutura secundária e formada de 5 folhas *Betas* e 4 laços, as interações eletrostáticas são geralmente favoráveis para Bc-Csp, mas desfavoráveis para Bs-CspB (MUELLER *et al.*, 2000). Nas hipertermofílica *Thermotoga marítima* (código PDB: 1G6P) apresenta 66 resíduos, contém 4 folhas Beta e 4 laços combinados em duas folhas beta antiparalelas que compõem uma estrutura de barril beta, em que as cadeias beta 1-4 são organizadas em uma topologia de chave grega (KREMER *et al.*, 2001).

Na figura 19 pode-se observar a estrutura das quatro proteínas Cold Shock homologas estudadas, na topologia de chave grega obtida através do software VMD (HUMPHREY; DALKE; SCHULTEN, 1996).

Figura 19 – Estrutura 3D das Proteínas Cold Shock na forma de fechadura grega.



As estruturas tridimensionais de proteínas Cold Shock homólogas apresentam basicamente a mesma quantidade de resíduos que são 66, com exceção para a mesofílica *Bacillus caldolyticus* que apresenta 67.

Durante o alinhamento de duas sequências proteicas devemos deixá-las similares ajustando as posições dos resíduos, se necessário, usando espaços (*gaps*), buscando assim aumentar a identidade entre duas proteínas, que é a porcentagem de resíduos idênticos presente em uma proteína em comparação com a outra (Figura 20).

Apesar dessa igualdade, a termoestabilidade está relacionada com as variações de posições que estes resíduos apresentam na sequência de cada proteína. Em geral, os resíduos localizados no interior das proteínas variam com menor frequência, e quando o fazem, ocorrem, normalmente, com menor distinção de propriedades físico-químicas. Habitualmente, certo conjunto de resíduos de aminoácidos que compreendem o núcleo da proteína e os principais elementos de estrutura secundária permanece mais conservado dentro de uma família de proteínas homólogas.

As maiores divergências entre proteínas homólogas ocorrem com mais frequência em regiões próximas da superfície, ou seja, nas rotações das ligações que formam a proteína, sendo que a estrutura secundária não está ainda definida. Nessas regiões, até mesmo as propriedades físico-químicas dos resíduos que sofreram mutações são muito diferentes dos resíduos anteriores ao processo de mutação (HÖLTJE *et al.*, 2003).

Uma das maneiras mais eficientes para se realizar uma busca de proteínas homólogas é através de similaridade sequencial, lançando-se mão de técnicas de bioinformática, uma vez que sequências de aminoácidos de proteínas com estruturas não-resolvidas podem ser obtidas com certa facilidade. Nesse caso, o grau de identidade sequencial, obtido pelo alinhamento com relação a uma ou várias sequências de estruturas conhecidas tridimensionalmente, e a predição das estruturas secundárias que os aminoácidos da sequência-alvo assumirão são aspectos primordiais na construção do modelo molecular da proteína-alvo (MUNIZ, 2003).

Nessa busca, que funciona como um processo de triagem, são utilizadas certas ferramentas para a avaliação do grau de similaridade entre fragmentos das sequências (alinhamento local), com o objetivo de distinguir entre similaridades importantes do ponto de vista biológico ou estrutural de similaridades ao acaso e que não apresentam importância significativa. A busca é realizada em um banco de dados com estruturas resolvidas, como o PDB. Com a introdução da sequência da proteína-alvo, um algoritmo de busca de similaridade de sequências é, então, aplicado.

Com respeito à faixa aceitável de identidade sequencial para a execução da modelagem molecular por homologia, é bem conhecido e a literatura descreve como significativo um valor acima de 30% de identidade sequencial entre a(s) proteína(s)-molde e a proteína-alvo (D'ALFONSO, et al., 2001; SALI, 1998; VITKUP, et al., 2001).

A figura 20 demonstra o alinhamento das proteínas Cold Shock realizado através da ferramenta computacional T-Coffee (NOTREDAME; HIGGINS; HERINGA, 2000), as sequências de resíduos de aminoácidos foram obtidas por meio do repositório de proteínas do site PDB (Protein Data Bank) através do arquivo de topologia FASTA.

Figura 20 – Alinhamento dos aminoácidos das proteínas homologas (Psicrofílica, Mesofílica, Termofílica e Hipertermofílica).

```

Psicrofílica  MEQGTVKWFNAEKGFIERENGDDVVFVHFSAIQGDGFKSLDEGQAVTFDVEEGQRGPQAANVQK-A
Mesofílica   MLEGKVKWFNSEKGFIEVEGQDDVVFVHFSAIQGEKFKTLEEGQAVSFEIVEGNRGPQAANVTKEA
Termofílica  MQRGKVKWFNNEKGYGFIEVEGGSDVVFVHFTAIQGEKFKTLEEGQEVSEIVEVQGNRGPQAANVVK-L
Hipertermofílica M-RGKVKWFDKKGYGFIKDEGGDVVFVHWSAIEMEGFKTLKEGQVVEFEIQEGKKGPQAAHVKVVE

```

Nas cores iguais são apresentados os aminoácidos semelhantes em duas, três ou quatro proteínas, para um melhor alinhamento utilizamos o hífen, visto que neste alinhamento só foi necessário o uso de 3 hifens, visto a grande quantidade de aminoácidos semelhantes.

Uma das maneiras frequentemente empregadas de se medir a similaridade entre duas proteínas é pelo cálculo do desvio da raiz quadrática média, ou do inglês, RMSD (Root-Mean-Square Deviation), que consiste na medida da distância média entre os átomos de proteínas sobrepostas.

Estruturalmente, as diferenças entre as proteínas podem ser caracterizadas pelo desvio médio da raiz quadrado médio, RMSD (NEUDERT; KLEBE, 2011). A tabela 3 apresenta os valores de RMSD das proteínas Cold Shock. Que mede as distâncias interatômicas (em Angstroms), entre o modelo gerado e a estrutura molde.

$$RMSD = (v.w) = \sqrt{\frac{1}{N} \sum_{i=1}^n \|v_i - w_i\|^2}$$

(9)

$$= \sqrt{\frac{1}{N} \sum_{i=1}^n (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$$

em que v e w são dois vetores que representam as posições dos átomos de cada sequência.

Tabela 3 – Determinação do RMSD das proteínas Cold Shock Homologas.

RMS	Psicrofílica (2LXJ)	Mesofílica (1CSP)	Termofílica (1C9O)	Hipertermofílica (1G6P)
Psicrofílica (2LXJ)	-	1.394	1.412	1.889
Mesofílica (1CSP)	1.394	-	0.369	2.029
Termofílica (1C9O)	1.412	0.369	-	2.143
Hipertermofílica (1G6P)	1.889	2.029	2.143	-

Fonte: Autor.

Também pode-se determinar a identidade entre duas proteínas sendo uma proteína-alvo e a outra a proteína molde, para isso após o alinhamento das proteínas, identificamos quantos resíduos estão presentes nas duas e fazemos a divisão pela quantidade de resíduos total, multiplicando esse valor por 100 temos a sua porcentagem.

Tabela 4 – Comparativo da identidade das proteínas Cold Shock em porcentagem.

IDENTIDADE	Psicrofílica (2LXJ)	Mesofílica (1CSP)	Termofílica (1C9O)	Hipertermofílica (1G6P)
Psicrofílica (2LXJ)	-	74,6	61,2	53,7
Mesofílica (1CSP)	74,6	-	82,1	62,7
Termofílica (1C9O)	61,2	82,1	-	62,7
Hipertermofílica (1G6P)	53,7	62,7	62,7	-

Fonte: Autor.

3. Materiais e Métodos.

3.1. Modelo Carbono- α .

Neste trabalho foram realizadas simulações computacionais de dinâmica molecular utilizando o modelo baseado em estrutura, tomando como base, o modelo C α (NYMEYER; GARCÍA; ONUCHIC, 1998; CLEMENTI; NYMEYER; ONUCHIC, 2000; KOGA; TAKADA, 2001; CHAVEZ; ONUCHIC; CLEMENTI, 2004; NOEL *et al.*, 2010; GOSAVI *et al.*, 2006). Este modelo diminui o tempo computacional devido a duas aproximações feitas; a primeira refere à cadeia polipeptídica onde serão considerados apenas os carbonos α dos resíduos dos aminoácidos e a segunda aproximação refere-se ao potencial de interação entre os carbonos α . Esse potencial é construído a partir da estrutura nativa da proteína em estudo. Fornecendo valores muito próximos dos que são obtidos, quando utilizados todos os carbonos e suas interações com o meio reacional.

As informações necessárias com as características das proteínas em estudo foram obtidas no PDB (*Protein Data Bank*) (BERMAN *et al.*, 2000); “RCSB PDB”. As estruturas obtidas do banco de dados de proteínas são chamadas de estado nativo para a elaboração do modelo em estudo. Por isso atribuímos a esse estado o mínimo de energia. O potencial para uma conformação de qualquer Γ de uma proteína, com base na conformação nativa Γ_0 é dado por:

$$\begin{aligned}
V(\Gamma, \Gamma_0) = & \sum_{\text{ligações}} E_r (r - r_0)^2 + \sum_{\text{ângulos}} E_\theta (\theta - \theta_0)^2 \\
& + \sum_{\text{cadeia principal}} E_{BB} \left\{ [(1 - \cos(\varphi - \varphi_0))] + \frac{1}{2} [1 - \cos(3(\varphi - \varphi_0))] \right\} \quad (10) \\
& + \sum_{\text{contatos}} E_c \left[5 \left(\frac{d_{ij}}{r_{ij}} \right)^{12} - 6 \left(\frac{d_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{\text{não contatos}} E_{NC} \left(\frac{\sigma_{NC}}{r_{ij}} \right)^{12}
\end{aligned}$$

Nesta expressão do potencial do modelo $C\alpha$, o primeiro termo representa a ligação entre dois carbonos α adjacentes, formando um potencial harmônico, Γ_0 onde é a distância entre dois carbonos α da conformação nativa ligados entre si. O segundo termo se refere ao potencial harmônico angular formado por três carbonos α consecutivos, onde Θ_0 é o ângulo formado pelos resíduos i , $i+1$ e $i+2$ da conformação nativa. O terceiro termo contabiliza a torção realizada na cadeia, o termo utiliza quatro carbonos α em sequência, os três primeiros carbonos α formam um plano que possui um ângulo Φ_0 com o plano formado entre três últimos carbonos α . O quarto termo representa a interação entre o carbono αi e o carbono αj que realizam um contato na estrutura nativa.

Para esta informação é atribuído um potencial 10 – 12, onde o parâmetro d_{ij} é o valor da distância entre estes carbonos que realizam um contato nativo.

O último termo é um potencial repulsivo utilizado para manter a distância máxima de aproximação entre os carbonos α , este termo é utilizado para todos os carbonos α que não possuem um contato nativo. O valor do parâmetro σ_{NC} de 4Å, este valor caracteriza aproximadamente o volume ocupado por um carbono do modelo. As constantes E_r e E_θ utilizadas no potencial possuem valores de 100, 20, 1 e 1, respectivamente, todas em unidades de E_c .

3.2. Simulação Computacional.

Os arquivos das proteínas foram obtidos do site RCSB PDB, topologia e coordenadas iniciais, foram criados pela interface gráfica do servidor SMOG-Server (NOEL; WHITFORD; ONUCHIC, 2012; NOEL *et al.* 2016; CLEMENTI; NYMEYER; ONUCHIC, 2000). As simulações computacionais foram realizadas com pacote LINUX (gratuitos) de dinâmica molecular GROMACS, versão 4.6.7 (HESS *et al.*, 2008; ASTUTI; MUTIARA, 2009), no início para utilizamos os computadores da UFTM, para o aprendizado, mais no intuito de acelerar os cálculos computacionais e diminuir o tempo computacional de espera

para os resultados foi utilizado o computador da UNESP – Gridunesp, onde apresenta supercomputadores dotados de vários núcleos, assim depois de realizada as simulações com os dados obtidos, foram feitos os cálculos para obtenção dos gráficos, utilizando os scripts e algoritmos criados pelo grupo e padronizado para cada situação.

Fez-se o estudo das trajetórias das proteínas Cold Shock Homologas, utilizando o algoritmo PYWHAM (FERRENBURG et al., 1988; KUMAR et al., 1992; HESS et al., 2008) para determinar o calor específico (CV) das proteínas em função da temperatura e a energia livre pelo número de contatos $F(Q)$ em função da temperatura.

3.2.1. Método Múltiplos Histogramas.

O método dos múltiplos histogramas ou WHAM (The Weighted Histogram Analysis Method) (FERRENBURG; SWENDSEN, 1988; FERRENBURG; SWENDSEN, 1989; KUMAR e at., 1992) é utilizado no cálculo de grandezas termodinâmicas. Esse método é uma generalização do método do histograma simples. No método do histograma simples fazemos a simulação em uma única temperatura e montamos um histograma de energia que mede a probabilidade de uma energia ocorrer em uma certa temperatura. Para sistemas maiores a simulação em uma única temperatura não é suficiente para varrer todo espaço conformacional. É preciso o uso de vários histogramas.

A função probabilidade para as temperaturas T e T' são dadas abaixo.

$$P_B(E, Q) = \frac{\Omega(E, Q)e^{-\beta E}}{\sum E\Omega(E, Q)e^{-\beta E}} \quad (11)$$

$$P_{B'}(E, Q) = \frac{\Omega(E, Q)e^{-\beta' E}}{\sum E\Omega(E, Q)e^{-\beta' E}} \quad (12)$$

onde β e β' são respectivamente $\frac{1}{k_B T}$ e $\frac{1}{k_B T'}$ e $\Omega(E, Q)$ é a densidade de estados obtidos através das simulações.

Fazendo se a divisão da equação 12 pela 11 obtemos

$$P_{B'}(E, Q) = \frac{P_B(E, Q)e^{-E(\beta' - \beta)}}{\frac{Z_{\beta'}}{Z_\beta}} \quad (13)$$

onde $Z_\beta = \sum E\Omega(E, Q)e^{-\beta E}$ e $Z_{\beta'} = \sum E\Omega(E, Q)e^{-\beta' E}$

A equação da probabilidade para uma temperatura T' é dada por:

$$P_{B'}(E, Q) = \frac{P_B(E, Q)e^{-E(\beta' - \beta)}}{\sum E P_B(E, Q)e^{-E(\beta' - \beta)}} \quad (14)$$

$$P_B(E, Q) = \frac{\Omega(E, Q)e^{-\beta E}}{\sum E \Omega(E, Q)e^{-\beta E}} \quad (15)$$

Para o cálculo de $P_B(E, Q)$ é necessário o cálculo da densidade de estados. Para uma simulação k .

$$\Omega_k(E, Q) = N_k(E, Q)e^{(\beta_k E - f_k)} \quad (16)$$

$$f_k = \beta_k A_k \quad (17)$$

Onde N_k é o valor tirado do histograma em E . A_k é a energia livre do sistema durante a simulação k .

$$A_k = \frac{\ln Z}{\beta_k} \quad (18)$$

Z é a função partição.

A densidade de estados $\Omega(E, Q)$ é dada por uma média ponderada das densidades de estados $\Omega_k(E, Q)$ ($k = 1, 2, 3, \dots, R$) obtidas por simulações.

$$\Omega(E, Q) = \sum_{j=1}^R w_j(E, Q) \Omega_k(E, Q) \quad (19)$$

$$\sum_{j=1}^R w_j(E, Q) = 1 \quad (20)$$

O conjunto de pesos que produz a melhor estimativa de $\Omega(E, Q)$ é obtido minimizando o erro estatístico $\delta^2 \Omega(E, Q)$. Para minimizar o erro apresentado por Ferrenberg and Swendsen (SWENDSEN, 1993), as derivadas parciais $X \frac{[e^{2\Omega(E, Q)}]}{X w_i}$ ($i = 1, 2, 3, \dots, R$) são igualadas a zero, resultando nas equações do método do múltiplo histograma.

$$P_B(E, Q) = \frac{\sum_{j=1}^R N_j(E, Q)e^{-\beta E}}{\sum_{j=1}^R n_j(E, Q)e^{(f_j - \beta E)}} \quad (21)$$

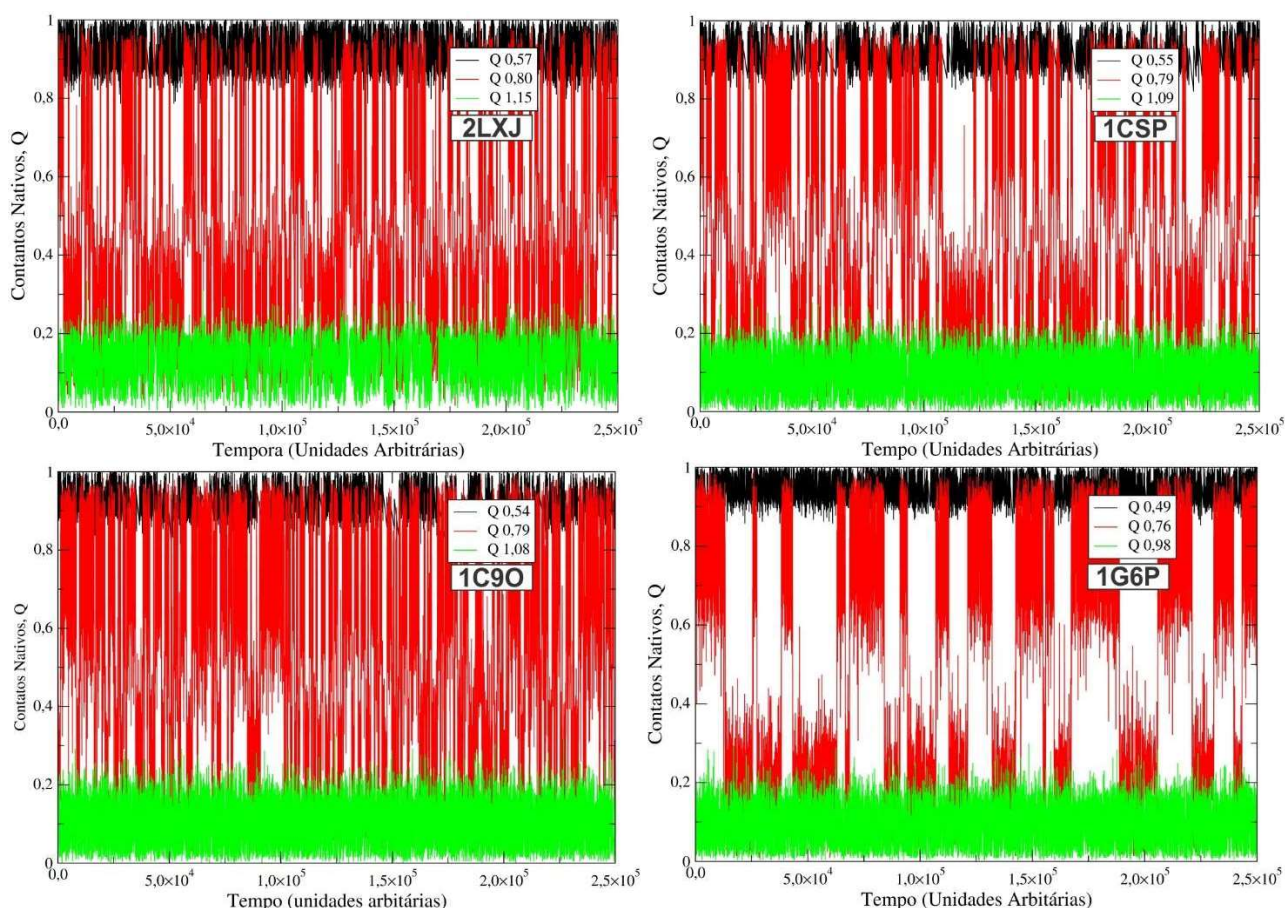
onde

$$e^{-f_j} = \sum_E P_{\beta_j}(E, Q) \quad (22)$$

4. Resultados e Discussão.

Ao fazer as simulações computacionais com as proteínas em estudo, foram observadas as dependências com a temperatura em que era realizada a simulação nas proteínas Cold Shock (Figura 19). Mesmo apresentando vários grupos similares entre as proteínas (Figura 20), as mesmas apresentaram pontos de T_f distinto entre elas, sendo notado esta similaridade no alinhamento realizado entre as proteínas.

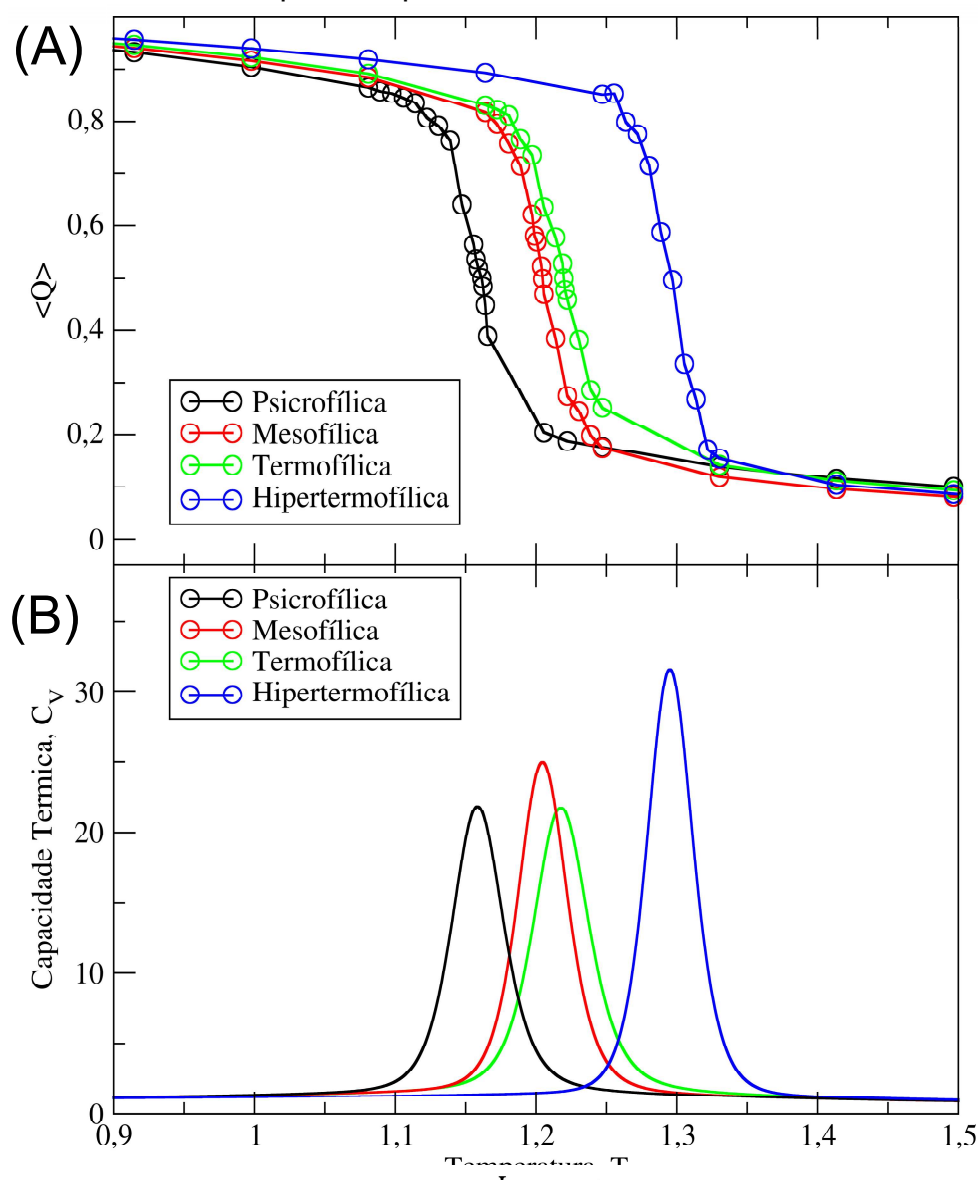
Figura 21 – Contatos nativos (Q) em função do tempo.



A figura 21 dos Contatos nativos (Q) em função do tempo computacional em unidades reduzidas do Gromacs: Neste gráfico, o qual foi normalizado para mais fácil visualização, pode-se observar a fração dos contatos nativos de uma proteína em função do tempo, a região verde está acima da temperatura de enovelamento, onde os contatos nativos são mínimos, ou seja, a proteína está desenovelada, na parte preta do gráfico corresponde à temperatura abaixo, onde os contatos nativos são máximos, ou seja a proteína se encontra enovelada, já a região vermelha do gráfico corresponde ao T_f de cada proteína. Nessa temperatura a fração de contatos nativos varia do seu mínimo para seu máximo, ficando em um estado de transição, se enovelando e desenovelando.

Foi calculado a capacidade térmica (Figura 22-B) em que se nota a variação de valor para cada uma das proteínas em estudo, demonstrando assim que a variação de apenas parte dos aminoácidos de uma cadeia pode causar alterações de suas propriedades, onde a Cold Shock Hipertermofílica apresentou a maior T_f , sendo por volta de 1.297 (unidades reduzidas). Os picos das capacidades térmicas demonstram também que estão ocorrendo variações devido aos grupos diferentes de aminoácidos entre cada uma das proteínas Cold Shock Homologas. Os valores de temperatura proteínas Cold Shock através da simulação computacional, corroboram com os dados experimentais (Tabela 4).

Figura 22 – Fração nativa em função das médias dos contatos nativos ($\langle Q \rangle$) e Capacidade Térmica pela temperatura em unidades reduzidas.



Fração nativa pela temperatura: Este gráfico, o qual foi normalizado para facilitar a comparação entre as proteínas homólogas, apresenta a taxa de variação de número de contatos nativos em função da

temperatura, assim pode-se observar que conforme a temperatura aumenta temos uma queda sutil no número de contatos nativos até que a proteína alcance a T_f , ou seja, em seu estado de transição. Nesse momento a queda do número de contatos nativos é brusca até que a proteína se desnatura completamente fazendo com que a curva fica praticamente retilínea com variações de contatos nativos muito pequenos, assim os contatos nativos ficam mínimos. Neste gráfico pode-se observar que a T_f temperatura aumenta gradativamente para cada proteína.

Tabela 5 – Temperatura de enovelamento teórico e experimental

Proteína	T_f	T_m (°C)
Psicrofílica (2LXJ)	1,156	40
Mesofílica(1CSP)	1,205	52
Termofílica (1C9O)	1,222	77
Hipertermofílica (1G6P)	1,297	80

Na tabela 4, temos os valores da temperatura de transição em unidades reduzidas bem como a temperatura experimental (LEE *et al.*, 2013), aqui também chamada de T_f , é considerada a temperatura de equilíbrio na qual metade das proteínas permanece na sua forma nativa, e a outra metade das proteínas está desnaturada. Do valor de T_f , pode-se avaliar a estabilidade térmica da proteína, pois quanto maior for a temperatura necessária para atingir um estado de entropia preponderante às forças químicas de estabilização, maior será a sua estabilidade na conformação nativa.

Na reação de enovelamento de proteínas é empregado a coordenada de reação Q , que corresponde a fração de contatos nativos formados na estrutura. Para monitorar a termodinâmica do sistema, deve-se agrupar as configurações obtidas durante uma simulação como uma função de Q . Sendo assim, o parâmetro Q representa a quantidade de contatos correspondentes na forma nativa, demonstrando-se como uma boa medida do grau de semelhança com a estrutura nativa (CLEMENTI; NYMEYER; ONUCHIC, 2000).

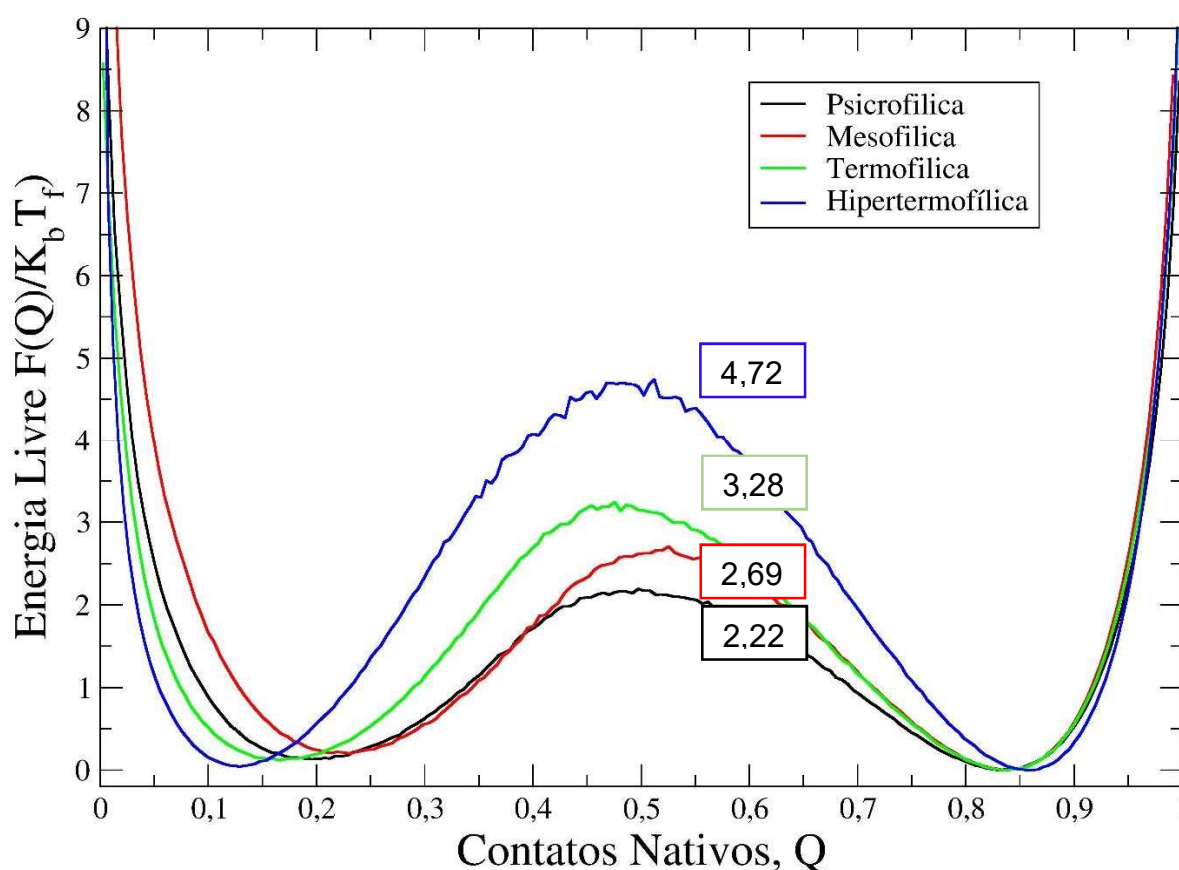
Nota-se que, por meio do perfil de energia livre em função de Q , se torna de fácil localização os agrupamentos desenovelados, enovelados e o estado de transição, como é mostrado na Figura 23. Por estes modelos considerarem as sequências não-frustradas, eles podem não gerar as mesmas energias das proteínas reais, assim como o valor das alturas de barreira e a estabilidade dos estados intermediários; porém há uma ótima concordância com resultados experimentais (CLEMENTI; NYMEYER; ONUCHIC, 2000).

As energias livres foram medidas em unidades de $k_b T_f$ a qual é denominada temperatura de transição do estado desenovelado para enovelado. Para um curto tempo de

enovelamento, mesmo apresentado uma barreira termodinâmica baixa, a energia térmica k_bT sobre a sua velocidade não pode ser desprezada.

Essas variações em energia livre nos estados nativo desenovelamento e enovelamento são evidenciadas na figura 23, onde os perfis de energia livre em função dos contatos nativos (Q) na temperatura de enovelamento (T_f) obtidos com as simulações em modelo baseados em C- α . Onde temos a transição do estado enovelado para o desenovelado, quando a proteína tem energia suficiente para romper esta barreira de energia livre, trocando assim de estado desenovelado para o estado enovelamento de menor energia.

Figura 23 – Energia livre em função dos contatos nativos (Q) na temperatura de enovelamento T_f .



Os perfis de energia livre em função da coordenada de reação fração de contatos nativos (Q) na temperatura de enovelamento (T_f) obtidos com as simulações em modelo C- α para as proteínas. Utilizou-se T_f como parâmetro definido através da temperatura no pico de calor específico de cada proteína simulada. Essas são proteínas com nenhuma ou baixíssima barreira de energia livre no estado de transição.

5. Conclusão.

Este trabalho procurou compreender a termoestabilidade das proteínas Cold Shock homologas avaliando assim as suas estabilidades frente a variações de temperatura, sendo empregados para isso modelos simplificados de simulações.

Conclui-se que por meio de estudos teóricos, é possível discutir um pouco das ideias que norteiam a busca pela compreensão da termoestabilidade das proteínas, com base em conceitos descrito na literatura sobre as proteínas homologas (SANTOS FILHO; ALENCASTRO, 2003; SILVA, 2007). Também pode-se destacar os aspectos gerais que são realizados em uma simulação computacional, que foi utilizado para as proteínas Cold Shock homologas (PERL *et al.*, 1998; PHADTARE; ALSINA; INOUE, 1999; ULUSU; TEZCAN, 2001), que foram Proteína Cold Shock da bactéria *Bacillus subtilis* (Bs-CspB), Proteína Cold Shock da bactéria *Bacillus caldolyticus* (Bc-Csp), Proteína Cold Shock da bactéria *Listeria monocytogenes* (Lm-Csp) e Proteína Cold Shock da bactéria *Thermotoga maritima* (Tm-Csp).

A analogia apresentada é uma aceitável confirmação do modelo usado na simulação computacional corroborando com o experimental, onde a proteína que apresenta maior estabilidade termodinâmica foi a Proteína Cold Shock da bactéria hipertermofílica *Thermotoga maritima* (Tm-Csp), confirmado pelos dados experimentais (LEE *et al.*, 2013). Sendo assim, o modelo baseado em estrutura, e a utilização apenas dos carbonos- α mostrou-se bem eficiente para investigar proteínas *in silico*.

Obtendo assim uma correlação entre o modelo apresentado neste trabalho e a simulação tanto experimental e computacional convergiram em grande acordo.

Outra variável de extrema importância foi a dos contatos nativos pelo tempo computacional onde obtivemos valores diferenciados entre as quatro proteínas em estudo.

Por fim, todos os objetivos propostos no início deste trabalho foram alcançados, além de que os resultados aqui obtidos apresentam grande relevância para área de estudo.

Para possíveis trabalhos futuros poderíamos verificar uma ou mais alterações na sequência de AA das proteínas, onde alteraria seus mecanismos de enovelamento e desenovelamento, aumentando ou diminuindo assim a sua temperatura de enovelamento, como os seus contatos nativos, dependendo da possível aplicação dessas proteínas.

6. Referências.

- ALBER, Tom. Stabilization energies of protein conformation. In: **Prediction of protein structure and the principles of protein conformation**. Springer, Boston, MA, 1989. p. 161-192.
- ALMEIDA, A. B. **Predição de estrutura terciária de proteínas com técnicas multiobjetivo no algoritmo de Monte Carlo**. 2016. 129 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás, Goiânia, 2016.
- ANFINSEN, Christian B. Principles that govern the folding of protein chains. **Science**, v. 181, n. 4096, p. 223-230, 1973.
- ASTUTI, A. D.; MUTIARA, A. B. Performance Analysis on Molecular Dynamics Simulation of Protein Using GROMACS. **arXiv preprint arXiv: 0912.0893**, 2009.
- BAYNES, Helton Godwin. **Mythology of the Soul (Psychology Revivals): A Research Into the Unconscious from Schizophrenic Dreams and Drawings**. Routledge, 2015.
- BENSON, D. A., CAVANAUGH, M., CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J., SAYERS, E. W. **GenBank. Nucleic acids research**, v. 41, n. D1, p. D36-D42, 2012.
- BERMAN, H. M., WESTBROOK, J., FENG, Z., GILLILAND, G., Bhat, T. N., WEISSIG, H., BOURNE, P. E. **The protein data bank. Nucleic acids research**, v. 28, n. 1, p. 235-242, 2000.
- CAMARGO, S. R. **Identificação de genes e uso de promotores modulados por etanol em cana-de-açúcar**. 2007. 84p. Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Biologia, Campinas, SP.
- CHANG, RAYMOND. **Physical chemistry for the chemical and biological sciences**. University Science Books, 2000.
- CHAVEZ, L. L.; ONUCHIC, J. N.; CLEMENTI, C. Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. **Journal of the American Chemical Society**, v. 126, p. 8426-8432, 2004.

CLEMENTI, C.; NYMEYER, H.; ONUCHIC, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. **Journal of molecular biology**, v. 298, p. 937-953, 2000.

CONTESSOTO, V. G., OLIVEIRA, V. M., CARVALHO, S. J., OLIVEIRA, L. C., LEITE, V. B. NTL9 folding at constant pH: the importance of electrostatic interaction and pH dependence. **Journal of chemical theory and computation**, v. 12, n. 7, p. 3270-3277, 2016.

CONTESSOTO, V. G., OLIVEIRA, V. M., FERNANDES, B. R., SLADE, G. G., LEITE, V. B. TKSA-MC: A Web Server for rational mutation through the optimization of protein charge interactions. **bioRxiv**, p. 22156, 2017.

CORNELL, W. D., CIEPLAK, P., BAYLY, C. I., GOULD, I. R., MERZ, K. M., FERGUSON, D. M., KOLLMAN, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. **Journal of the American Chemical Society**, v. 117, n. 19, p. 5179-5197, 1995.

D'ALFONSO, G.; TRAMONTANO, A.; LAHM, A. Structural conservation in single-domain proteins: implications for homology modeling. *Journal of Structural Biology*. v. 134, p. 246-256, 2001.

DA SILVA, VINICIUS B. Modelagem molecular de proteínas-alvo por homologia estrutural. **Revista eletrônica de Farmácia**, v. 4, n. 1, 2007.

DE GODOI CONTESSOTO, V., DE OLIVEIRA JUNIOR, A. B., CHAHINE, J., DE OLIVEIRA, R. J., LEITE, V. B. P. (2018) Introdução ao problema de enovelamento de proteínas: uma abordagem utilizando modelos computacionais simplificados. **Revista Brasileira de Ensino de Física**, v. 40, n. 4, p. e4307, 2018.

DILL, K.; BROMBERG, S. Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience. [S.l.]: **Garland Science**, 2012.

DOBSON, C. M. Protein folding and misfolding. **Nature**, v. 426, p. 884, 2003.

DOYE, J. P. K.; MILLER, M. A.; WALES, D. J. Evolution of the potential energy surface with size for Lennard-Jones clusters. **The Journal of chemical physics**, v. 111, p. 8417-8428, 1999 - a.

DOYE, J. P. K.; MILLER, M. A.; WALES, D. J. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. **The Journal of Chemical Physics**, v. 110, p. 6896-6906, 1999 - b.

DURHAM, E. H. A. B. **Bioinformática estrutural de proteínas modificadas por eventos de splicing alternativo**. 2007. 131 f. Tese de Doutorado. Universidade de São Paulo, São Paulo, 2007.

EVANS, DAVID A.; WALES, DAVID J. The free energy landscape and dynamics of met-enkephalin. **The Journal of chemical physics**, v. 119, n. 18, p. 9947-9955, 2003.

FERRENBURG, ALAN M.; SWENDSEN, ROBERT H. New Monte Carlo technique for studying phase transitions. **Physical review letters**, v. 61, n. 23, p. 2635, 1988.

FERRENBURG, A. M.; SWENDSEN, R. H. Optimized monte carlo data analysis. **Computers in Physics**, v. 3, p. 101-104, 1989.

FISCHER, DANIEL. Servers for protein structure prediction. **Current opinion in structural biology**, v. 16, n. 2, p. 178-182, 2006.

GONZALEZ, FELIX H. DIAZ; DA SILVA, SERGIO CERONI. Introdução à bioquímica clínica veterinária. 2017.

GOSAVI, S., CHAVEZ, L. L., JENNINGS, P. A., ONUCHIC, J. N. Topological frustration and the folding of interleukin-1 β . **Journal of molecular biology**, v. 357, n. 3, p. 986-996, 2006.

HARVEY, R. A.; FERRIER, D. R. **Bioquímica ilustrada**. [S.l.]: Artmed Editora, 2015.

HESS B, KUTZNER C, VAN DER SPOEL D, LINDAHL E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. **Journal of Chemical Theory and Computation**. 2008;4(3):435-47.

HÖLTJE, H. D.; SIPPL, W.; ROGNAN, D.; FOLKERS, G. Introduction to comparative protein modeling. In **Molecular Modeling: Basic Principles and Applications**. Weinheim: Wiley-VCH, 2003.

HUMPHREY, WILLIAM; DALKE, ANDREW; SCHULTEN, KLAUS. VMD: visual molecular dynamics. **Journal of molecular graphics**, v. 14, n. 1, p. 33-38, 1996.

ITZHAKI, L. S.; OTZEN, D. E.; FERSHT, A. R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. **Journal of molecular biology**, v. 254, p. 260-288, 1995.

KENDREW, J. C., BODO, G., DINTZIS, H. M., PARRISH, R. G., WYCKOFF, H., PHILLIPS, D. C. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. **Nature**, v. 181, n. 4610, p. 662-666, 1958.

KOGA, N.; TAKADA, S. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. Proceedings of the National Academy of Sciences, 2012. **Journal of molecular biology**, v. 313, p. 171-180, 2001.

KREMER, W., SCHULER, B., HARRIEDER, S., GEYER, M., GRONWALD, W., WELKER, C., KALBITZER, H. R. Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. **European journal of biochemistry**, v. 268, n. 9, p. 2527-2539, 2001.

KRIVOV, S. V.; KARPLUS, M. Free energy disconnectivity graphs: Application to peptide models. **The Journal of chemical physics**, v. 117, p. 10894-10903, 2002.

KUMAR, S., ROSENBERG, J. M., BOUZIDA, D., SWENDSEN, R. H., KOLLMAN, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. **Journal of Computational Chemistry**. 1992;13(8):1011-21.

LEE, J.; FREDDOLINO, P. L.; ZHANG, Y. Ab initio protein structure prediction. **From protein structure to function with bioinformatics**. Springer, 2017. p. 3-35.

LI, HAO; TANG, CHAO; WINGREEN, NED S. Nature of driving force for protein folding: a result from analyzing the statistical potential. **Physical review letters**, v. 79, n. 4, p. 765, 1997.

LESK, V. I.; STERNBERG, M. J. E. 3D-Garden: a system for modelling protein--protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. **Bioinformatics**, v. 24, p. 1137-1144, 2008.

LEVINTHAL, C. Are there pathways for protein folding? **Journal de chimie physique**, v. 65, p. 44-45, 1968.

LUCAS, TÂNIA GOMES. **Estudos de folding e efeito de chaperões químicos e moleculares em proteínas envolvidas na β -oxidação de ácidos gordos**. 2011. Tese de Doutorado. Faculdade de Ciências e Tecnologia.

MILLER, M. A.; DOYE, J. P. K.; WALES, D. J. Structural relaxation in atomic clusters: Master equation dynamics. **Physical Review E**, v. 60, p. 3701, 1999.

MILLER, STANLEY L. A production of amino acids under possible primitive earth conditions. **Science**, v. 117, n. 3046, p. 528-529, 1953.

MOREIRA, L. M., MORAES, P. C. G. D., MENDONÇA, J. P. R. F. D., GUIMARÃES, L., LYON, J. P., AIMBIRE, F., IMASATO, H. Hemoglobina extracelular gigante de *Glossoscolex paulistus*: um extraordinário sistema supramolecular hemoproteico. **Química Nova**, v. 34, n. 1, p. 119-130, 2011.

MUELLER, U., PERL, D., SCHMID, F. X., HEINEMANN, U. Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein1. **Journal of molecular biology**, v. 297, n. 4, p. 975-988, 2000.

MUNIZ, J. R. C. **Aplicação da bioinformática nos estudos dos genes e enzimas envolvidos na síntese da goma fastidiana produzida pela *Xylela fastidiosa***. 2003. 124p. Dissertação (Mestrado) Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2003.

NAKAMURA, H. Roles of electrostatic interaction in proteins. **Quarterly reviews of biophysics**, v. 29, p. 1-90, 1996.

NELSON, DAVID LEE; COX, MICHAEL M. **Princípios de bioquímica de Lehninger**. Porto Alegre: Artmed, 2014.

NELSON, D. L.; LEHNINGER, A. L.; COX, M. M. **Lehninger principles of biochemistry**. [S.I.]: Macmillan, 2008.

NEUDERT, G.; KLEBE, G. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein–Ligand Complexes. **Journal of Chemical Information and Modeling**, v. 51, p. 2731-2745, out. 2011.

NOEL, J. K., WHITFORD, P. C., SANBONMATSU, K. Y., ONUCHIC, J. SMOG@ ctbp: simplified deployment of structure-based models in GROMACS. **Nucleic acids research**, v. 38, p. W657--W661, 2010.

NOEL, JEFFREY K.; WHITFORD, PAUL C.; ONUCHIC, JOSÉ N. The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. **The Journal of Physical Chemistry B**, v. 116, n. 29, p. 8692-8702, 2012.

NOEL, J. K., LEVI, M., RAGHUNATHAN, M., LAMMERT, H., HAYES, R. L., ONUCHIC, J. N., WHITFORD, SMOG 2: A versatile software package for generating structure-based models. **PLoS computational biology**, v. 12, n. 3, p. e1004794, 2016.

NOTREDAME, CÉDRIC; HIGGINS, DESMOND G.; HERINGA, JAAP. T-coffee: a novel method for fast and accurate multiple sequence alignment¹. **Journal of molecular biology**, v. 302, n. 1, p. 205-217, 2000.

NYMEYER, H.; GARCÍA, A. E.; ONUCHIC, J. N. Folding funnels and frustration in off-lattice minimalist protein landscapes. **Proceedings of the National Academy of Sciences**, v. 95, p. 5921-5928, 1998.

OLIVEIRA, R. J. **Estudo do coeficiente de difusão no enovelamento de proteínas na rede**. 2007. 78 f. Dissertação (mestrado) - Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2007

OLIVEIRA, R. J. **Estudo do coeficiente de difusão no enovelamento de proteína**. 2011. 115 f. Tese (doutorado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto, 2011.

ONUCHIC, JOSÉ NELSON; LUTHEY-SCHULTEN, ZAIDA; WOLYNES, PETER G. Theory of protein folding: the energy landscape perspective. **Annual review of physical chemistry**, v. 48, n. 1, p. 545-600, 1997.

PAVLOPOULOU, A.; MICHALOPOULOS, I. State-of-the-art bioinformatics protein structure prediction tools. **International journal of molecular medicine**, v. 28, p. 295-310, 2011.

PERL, D., WELKER, C., SCHINDLER, T., SCHRÖDER, K., MARAHIEL, M. A., JAENICKE, R., & SCHMID, F. X. Conservation of rapid two-state folding in mesophilic, thermophilic and

hyperthermophilic cold shock proteins. **Nature Structural and Molecular Biology**, v. 5, n. 3, p. 229, 1998.

PHADTARE, SANGITA; ALSINA, JANIVETTE; INOUE, MASAYORI. Cold-shock response and cold-shock proteins. **Current opinion in microbiology**, v. 2, n. 2, p. 175-180, 1999.

POTOYAN, D. A.; PAPOIAN, G. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. **Journal of the American Chemical Society**, v. 133, p. 7405-7415, 2011.

PRIVALOV, P. L.; KHECHINASHVILI, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. **Journal of molecular biology**, v. 86, p. 665-684, 1974.

RAZVI, A.; SCHOLTZ, J. M. Lessons in stability from thermophilic proteins. **Protein Science**, v. 15, p. 1569-1578, 2006.

REDDY, A. S., WANG, L., SINGH, S., LING, Y. L., BUCHANAN, L., ZANNI, M. T., SKINNER, J. L., Pablo, J. J. Stable and metastable states of human amylin in solution. **Biophysical journal**, v. 99, p. 2208-2216, 2010.

RUSSO, S.; SILVER, M. **Essentials of Introductory Chemistry**. [S.l.]: Addison-Wesley Longman, Limited, 2001. ISBN ISBN: 9780321046383. Disponível em: <<https://books.google.com.br/books?id=vwJqgwfx4H0C>>.

SABOURY, A. A.; MOOSAVI-MOVAHEDI, A. A. Clarification of calorimetric and van't hoff enthalpies for evaluation of protein transition states. **Biochemical Education**, v. 22, n. 4, p. 210-211, 1994.

SALGUEIRO, M. E. F. **Simulações on-line para o ensino e aprendizagem de Química**, 2003.

SALI, A. 100,000 protein structures for the biologist. **Nature Structural & Molecular Biology**. v. 05, p. 1029-1032, 1998.

SANTOS FILHO, OSVALDO ANDRADE; ALENCASTRO, RB de. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253-259, 2003.

SCHINDELIN, HERMANN; MARAHIEL, MOHAMED A.; HEINEMANN, UDO. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. **Nature**, v. 364, n. 6433, p. 164, 1993.

SHAN, Y., ARKHIPOV, A., KIM, E. T., PAN, A. C., SHAW, D. E. Transitions to catalytically inactive conformations in EGFR kinase. **Proceedings of the National Academy of Sciences**, p. 201220843, 2013.

SILVA, R. I. **Enovelamento proteico: fatores topológicos**. 2005. 74 p. Dissertação (Doutorado em Ciências) - Faculdade de Filosofia Ciências e Letras, Universidade de São Paulo, Ribeirão Preto.

STERPONE, FABIO; MELCHIONNA, SIMONE. Thermophilic proteins: insight and perspective from in silico experiments. **Chemical Society Reviews**, v. 41, n. 5, p. 1665-1676, 2012.

SWENDSEN, R. H. Modern methods of analyzing Monte Carlo computer simulations. **Physica A: Statistical Mechanics and its Applications**, v. 194, p. 53-62, 1993.

TEIXEIRA, Viviane Gomes; COUTINHO, Fernanda MB; GOMES, Ailton S. Principais métodos de caracterização da porosidade de resinas à base de divinilbenzeno. **Química nova**, v. 24, n. 6, p. 808-818, 2001.

ULUSU, N. NURAY; TEZCAN, E. FERHAN. Cold shock proteins. **Turkish Journal of Medical Sciences**, v. 31, n. 4, p. 283-290, 2001.

VITKUP, D.; MELAMUD, E.; MOULT, J.; SANDER, C. Completeness in structural genomics. **Nature Structural & Molecular Biology**. v. 08, p. 559-566, 2001.

WALES, D. **Energy landscapes: Applications to clusters, biomolecules and glasses**. [S.l.]: Cambridge University Press, 2003.

WALES, D. J. Energy landscapes: some new horizons. **Current opinion in structural biology**, v. 20, p. 3-10, 2010.

WALES, D. J. Decoding the energy landscape: extracting structure, dynamics and thermodynamics. **Phil. Trans. R. Soc. A**, v. 370, p. 2877-2899, 2012.

WANG, J., OLIVEIRA, R. J., CHU, X., WHITFORD, P. C., CHAHINE, J., HAN, W., LEITE, V. B. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. **Proceedings of the National Academy of Sciences**, 2012.

WELKER, C., PERL, D., SCHINDLER, T., SCHRÖDER, K., MARAHIEL, M. A., JAENICKE, R., SCHMID, F. X. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. **Nature Structural and Molecular Biology**, v. 5, n. 3, p. 229, 1998.

XIAO, L.; HONIG, B. Electrostatic contributions to the stability of hyperthermophilic proteins¹. **Journal of molecular biology**, v. 289, p. 1435-1444, 1999.

ZHURAVLEV, P. I.; PAPOIAN, G. A. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. **Quarterly reviews of biophysics**, v. 43, p. 295-332, 2010.

ZORDAN, A. B. **Estudo sobre o efeito da pressão em pectinases do suco de frutas utilizando método de simulação**. 2017. 102 f. Dissertação de Mestrado. Universidade Federal do Espírito Santo, Espírito Santo, 2017.