

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS GRADUAÇÃO EM PROCESSOS QUÍMICOS E  
BIOTECNOLÓGICOS**

**FABIO RODRIGUES**

**ESTUDO *IN SILICO* DA CO-FORMULAÇÃO DE INSULINA COM  
AMILINA E SEUS ANÁLOGOS BIOTECNOLÓGICOS**

**DISSERTAÇÃO DE MESTRADO  
PROGRAMA DE PÓS-GRADUAÇÃO EM PROCESSOS QUÍMICOS E  
BIOTECNOLÓGICOS  
PPGQB**

**TOLEDO  
2020**

**FABIO RODRIGUES**

**ESTUDO *IN SILICO* DA CO-FORMULAÇÃO DE INSULINA COM  
AMILINA E SEUS ANÁLOGOS BIOTECNOLÓGICOS**

Dissertação de mestrado apresentada como requisito parcial para obtenção do título de Mestre em Processos Químicos e Biotecnológicos, do Programa de Pós-Graduação em Processos Químicos e Biotecnológicos – PPGQB – da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Toledo.

Orientador: Prof. Dr. Rafael Bertolini Frigori

**TOLEDO  
2020**

Ficha catalográfica elaborada pela Bibliotecária Carla Rech Ribeiro - CRB 9/1685

R696e Rodrigues, Fábio

Estudo in silico da co-formulação de insulina com amilina e seus análogos biotecnológicos / Fábio Rodrigues; orientador: Rafael Bertolini Frigori. -- Toledo, PR: UTFPR, 2020.

67 f.: il.; 30 cm.

Texto em português

Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Processos Químicos e Biotecnológicos, Toledo, 2020.

1. Proteínas. 2. Termodinâmica. 3. Diabetes. 4. Frigori, Rafael Bertolini. I. Universidade Tecnológica Federal do Paraná - Programa de Pós-Graduação em Processos Químicos e Biotecnológicos. II. Título.

CDD: Ed. 22 -- 660.6



Ministério da Educação  
**Universidade Tecnológica Federal do Paraná**  
Câmpus Toledo  
PROGRAMA DE PÓS GRADUAÇÃO EM  
PROCESSOS  
QUÍMICOS E BIOTECNOLÓGICOS



---

## TERMO DE APROVAÇÃO

### **ESTUDO *IN SILICO* DA CO-FORMULAÇÃO DE INSULINA COM AMILINA E SEUS ANÁLOGOS BIOTECNOLÓGICOS**

por

**FABIO RODRIGUES**

Esta dissertação de mestrado foi apresentada em 07 de agosto de 2020 como requisito parcial para a obtenção do título de Mestre em Processos Químicos e Biotecnológicos. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

---

Prof. Dr. Rafael Bertolini Frigori  
(Orientador)

---

Prof.<sup>a</sup> Dra. Rosemeire Aparecida da Silva de Lucca  
(Membro Externo Titular)

---

Prof. Dr. Gilberto da Cunha Gonçalves  
(Membro Interno Titular)

- O Termo de Aprovação assinado encontra-se na Coordenação do Curso –

*Dedico este mestrado primeiramente e inteiramente à Deus por cada dia de minha vida, e também à minha esposa e família, pelos momentos de dificuldades.*

## **AGRADECIMENTOS**

A Deus por cada dia de minha vida, e por me dar força nos momentos difíceis dessa caminhada.

Ao meu orientador Prof. Dr. Rafael Bertolini Frigori, pela sabedoria e dedicação com que me guiou nesta trajetória.

Aos meus colegas de iniciação científica Luan e Leticia pela contribuição e compartilhamento de conhecimento.

A coordenação do curso, pela cooperação, e a todos os docentes pelo apoio e atenção.

Ao Laboratório Nacional de Computação Científica (LNCC) pelas horas CPUs cedidas ao nosso projeto;

A coordenação de ensino de Pós Graduação da CAPES pelo apoio e incentivo a pesquisa;

Gostaria de deixar registrado também, o meu reconhecimento à minha esposa e família, pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Enfim, a todos os que por algum motivo contribuíram para a realização deste estudo.

*“Existem muitas hipóteses em ciência que estão erradas. Isso é perfeitamente aceitável, elas são a abertura para achar as que estão certas” (Carl Sagan).*

## RESUMO

RODRIGUES, Fabio. **Estudo *in silico* da co-formulação de Insulina com Amilina e seus Análogos Biotecnológicos**: 2020. 67 folhas. Trabalho de conclusão de curso Mestrado em Processos Químicos e Biotecnológicos - Universidade Tecnológica Federal do Paraná. Toledo, 2020.

A diabetes é uma doença consequente da deficiência da produção de insulina ou resistência à sua ação. Na diabetes tipo **I**, o paciente já nasce com a doença hereditariamente, enquanto que no tipo **II**, sua gênese se dá como resultado de outros fatores e hábitos que contribuem para o surgimento dessa disfunção orgânica, como os de obesidade e sedentarismo, por exemplo, que são adquiridos pela má alimentação, provocada principalmente pela ingestão excessiva de carboidratos e gorduras, aliada a falta de atividades físicas. A doença é controlada através do acompanhamento do nível glicêmico e com aplicações de injeções de insulina, o que ocasiona um desconforto para o paciente, além de resultar em um ciclo de ganho de peso e resistência à insulina. A Amilina humana (hIAPP) é outro hormônio que atua na regulação metabólica junto com a insulina. No entanto devido ao seu perfil amiloidegênico, ou seja, agrega-se e forma fibrilas insolúveis, a hIAPP é inapropriada para uma co-formulação com insulina. Isso levou ao desenvolvimento de novos análogos da hIAPP com características melhoradas, através de técnicas de substituição de aminoácidos, como a Pramlintide conhecida como sIAPP, que é quantificada pela substituição de uma Alanina e duas Serinas, por Prolinas ou seja (i.e. A25P, S28P, S29P). Assim, sIAPP apresenta um significativo aumento de resistência a agregação. Mas resta, ainda, o problema de baixa solubilidade de sIAPP em pHs fisiológicos. Este problema foi estudado recentemente e levou à criação de sIAPP<sup>+</sup>: ou seja, uma substituição extra de uma Serina por uma Arginina na posição 20 (i.e. S20R) da sequência de sIAPP. Nesse trabalho, desenvolvemos um estudo teórico comparativo *in silico* empregando Amilina e seus análogos (i.e. hIAPP, sIAPP e sIAPP<sup>+</sup>) com intuito de testar a efetividade de suas coformulações com insulina. Em nossa metodologia, avaliamos a estabilidade dessas formulações por meio de dados extraídos da etapa microcanônica das simulações multicanônicas da coformulação de insulina com análogos de Amilina. Assim, calculamos grandezas termodinâmicas, como calores latentes e barreiras de energias-livres de agregação, avaliamos os resultados e verificamos teoricamente a estabilidade físico-química. Também analisamos as transições de fases e seus respectivos tempos de agregação. Nossos resultados são complementados com cálculos de energia de solvatação e análises estruturais.

**Palavras-chave:** Insulina. Amilina. Pramlintide. Análogos. Agregação.



## ABSTRACT

RODRIGUES, Fabio. **In silico study of the coformulation of Insulin with amylin and its biotechnological analogues**: 2020. 67 leaves. Course conclusion work Master in Chemical and Biotechnological Processes - Federal Technology University - Paraná. Toledo, 2020.

Diabetes is a disease resulting from a deficiency in insulin production or resistance to its action. In type one diabetes, the patient is already born with the disease hereditarily, while on type two, its genesis occurs as a result of other factors and habits that contribute to the appearance of this organic dysfunction, such as obesity and sedentary lifestyle, for example, which are acquired by poor diet, caused mainly by excessive intake of carbohydrates and fats, combined with a lack of physical activities. The disease is controlled by monitoring the glycemic level and using insulin injections, which causes discomfort for the patient, in addition to resulting in a cycle of weight gain and insulin resistance. Human amylin (hIAPP) is another hormone that acts on metabolic regulation along with insulin. However, due to its amyloidogenic profile, that is, it aggregates and forms insoluble fibrils, a hIAPP is inappropriate for co-use with insulin. This led to the development of new hIAPP analogs with improved characteristics, through amino acid substitution techniques, such as Pramlintide known as sIAPP, which is quantified by the replacement of an Alanine and two Serines, for Prolines ie (for example, A25P, S28P, S29P). Thus, sIAPP presents a significant increase in resistance to aggregation. But there is still the problem of low solubility of sIAPP at physiological pHs. This problem has been studied recently and led to the creation of the sIAPP<sup>+</sup>: that is, an extra replacement of a Serine with na Arginine at position 20 (i.e, S20R) of the sIAPP sequence. In this work, we developed a comparative theoretical study in silicon using Amylin and its analogues (i.e, hIAPP, sIAPP and sIAPP<sup>+</sup>) in order to test the effectiveness of its co-formulations with insulin. In our methodology, we evaluated the stability of these formulations using data extracted from the microcanonical stage of multicanonical simulations of insulin co-formulation with Amylin analogs. Thus, we calculate thermodynamic quantities, such as latent heat and free-energy aggregation barriers, evaluate the results and theoretically verify the physical-chemical stability. We also analyze the phase transitions and their respective aggregation times. Our results are complemented with energy calculations of solvation and structural analysis.

**Keywords:** Insulin. Amilin. Pramlintide. Analogues. Aggregation.

## LISTA DE ILUSTRAÇÕES

### FIGURAS

<b>Figura 1-</b> Fórmula molecular e abreviação dos 20 aminoácidos.....	08
<b>Figura 2-</b> Ligação covalente do dipeptídeo (fibroína).....	09
<b>Figura 3-</b> Exemplo de sequência de aminoácidos com 66 resíduos PDB 1BW5.....	15
<b>Figura 4-</b> Estrutura ( <i>backbone</i> ).....	16
<b>Figura 5-</b> Folhas betas e hélices alfa com ligações químicas.....	16
<b>Figura 6-</b> Mecanismo priônico.....	17
<b>Figura 7-</b> Formação de fibrilas.....	18
<b>Figura 8-</b> Domínio terciária de um tipo de insulina, PDB 1BW5 e Estrutura hemoglobina 3WCU respectivamente.....	18
<b>Figura 9-</b> Transição de primeira ordem.....	27
<b>Figura 10-</b> Curvas calórica ( $T \times \epsilon$ ) e da entropia microcanônica ( $S \text{ vs } \epsilon$ ) ao longo de uma transição de fase.....	28
<b>Figura 11-</b> Curvas calóricas para análogos de amilina em associação com insulina humana de ação regular.....	43
<b>Figura 12-</b> Mapa de calor hIAPP com Insulina.....	46
<b>Figura 13 –</b> Raio de giração ( $R_{gyr}$ ) e distância entre os terminais ( $R_{ee}$ ) da hIAPP com Insulina.....	46
<b>Figura 14 -</b> Mapa de calor sIAPP com Insulina.....	47
<b>Figura 15 –</b> Raio de giração ( $R_{gyr}$ ) e distância entre os terminais ( $R_{ee}$ ) da sIAPP com Insulina.....	47
<b>Figura 16 -</b> Mapa de Calor sIAPP <sup>+</sup> com Insulina.....	47

<b>Figura 17</b> – Raio de giração ( <i>R<sub>gyr</sub></i> ) e distância entre os terminais ( <i>Ree</i> ) da sIAPP <sup>+</sup> com Insulina.....	47
<b>Figura 18:</b> Agregação da hIAPP com insulina.....	48
<b>Figura 19:</b> Agregação da sIAPP com insulina.....	48
<b>Figura 20:</b> Agregação da sIAPP <sup>+</sup> com Insulina.....	48

## QUADROS

<b>Quadro 1:</b> Classificação dos aminoácidos, características químicas.....	07
---	----

## TABELAS

<b>Tabela 2:</b> Sequência de aminoácidos FASTA para Amilina humana, Pramlintide, S20R-Pramlintide com as respectivas mutações em vermelho e as cadeias de Insulina humana de ação regular (4EWZ).....	39
<b>Tabela 3:</b> Resultados termodinâmicos para a heterodimerização de análogos de amilina com a insulina humana de ação regular.....	44

## LISTA DE ACRÔNIMOS

AVC	Acidente Vascular Cerebral
CNT	Classical nucleation theory
FDA	Food and Drug Administration
IAPP	Islet Amyloid Polypeptide (amilina)
hIAPP	Human Islet Amyloid Polypeptide (amilina humana)
LNCC	Laboratório Nacional de Computação Científica
MUCA	Algoritmo Multicanônico
PDB	Protein Data Bank
PHAST	Protein-like Heteropolymer Analysis by Statistical Thermodynamics
siAPP	Pramlintide
siAPP <sup>+</sup>	Análogo da Pramlintide
S20R	Mutante da Pramlintide
SMMP	Simple Molecular Mechanics for Proteins
SSH	Secure Shell
ThT	Tioflavina

## LISTA DE SIMBOLOS

Å	Angstrom
$\epsilon$	Energia Interna por resíduo
S	Entropia
T	Temperatura
$T_c$	Temperatura Critica
$\Delta G$	Energia de Ativação
$\Delta L$	Calor Latente de Transição
$\Delta F$	Energia Livre de Helmholtz
G	Energia Livre de Gibbs
K	Escala Kelvin
$k_B$	Constante de Boltzmann
J	Joule
L	Calor Latente

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>1</b>
1.2 OBJETIVO GERAL.....	4
1.3 OBJETIVO ESPECIFICOS.....	4
<b>2 PROTEÍNAS: UMA BREVE REVISÃO BIBLIOGRÁFICA</b> .....	<b>5</b>
2.1 ESTUDO DOS AMINOÁCIDOS, PEPTÍDIOS E PROTEÍNAS.....	5
2.2 CARACTERIZAÇÃO QUÍMICA DOS AMINOÁCIDOS.....	6
2.3 O ENOVELAMENTO E DOBRAMENTO PROTEÍNAS.....	10
2.4 FUNCIONALIDADES DAS PROTEÍNAS.....	12
2.4.1 CLASSIFICAÇÃO DAS PROTEÍNAS.....	12
2.4.1.2 Classificação Morfológica das Proteínas.....	13
2.4.2 REPRESENTAÇÕES COMPUTACIONAIS DAS PROTEÍNAS.....	13
2.4.3 HIERARQUIA ESTRUTURAL DAS PROTEÍNAS.....	14
2.4.3.1 Estrutura Primária.....	14
2.4.3.2 Estrutura Secundária.....	15
2.4.3.3 Estrutura Terciária e quartenária.....	18
2.5 PROPRIEDADES FÍSICO-QUÍMICAS DAS PROTEÍNAS E AS INTERAÇÕES MOLECULARES.....	19
<b>3 PROPRIEDADES DA TERMODINÂMICA DE EQUILÍBRIO</b> .....	<b>22</b>
<b>4 O PRINCÍPIO DE TERMOESTATÍSTICA MICROCANÔNICA</b> .....	<b>24</b>
4.1 O ENSEMBLE CANÔNICO.....	25
4.2 O ENSEMBLE MICROCANÔNICO.....	26
4.3 TRANSIÇÕES DE FASES TERMODINÂMICA.....	26
4.4 TEORIA CLÁSSICA DE NUCLEAÇÃO .....	28
<b>5 SIMULAÇÕES COMPUTACIONAIS</b> .....	<b>30</b>
5.1 MODELOS DE SIMULAÇÕES.....	30
5.1.2 Simulações de Monte Carlo.....	30
5.2 CAMPO DE FORÇA ECEPP/3.....	32
5.2.1 Modelos de Solvatação.....	34
5.2.2 Modelagem da Solvatação Com o PBEQ-Solver.....	35
5.3 SOFTWARES DE VISUALIZAÇÃO E MODELAGEM MOLECULAR.....	36
<b>6 MATERIAIS E MÉTODOS</b> .....	<b>38</b>
<b>7 RESULTADOS E DISCUSSÕES</b> .....	<b>43</b>
<b>8 CONCLUSÕES</b> .....	<b>49</b>
<b>9 REFERÊNCIAS</b> .....	<b>50</b>
<b>APÊNDICE A – RAIOS DE GIRAÇÃO</b> .....	<b>55</b>

## 1 INTRODUÇÃO

A diabetes melito do tipo **I** é uma doença degenerativa e congênita, provocada pela deficiência da secreção de insulina pelo pâncreas, em decorrência da destruição das células pancreáticas. Já a diabetes melito tipo **II**, pode ser causada em decorrência de alguns agravantes como obesidade, sedentarismo e fatores genéticos que resultam em uma certa peculiaridade: a “resistência à insulina”, fazendo com que o paciente precise progressivamente de mais doses de insulina (MELMED, 2001). A insulina é um hormônio anabólico dimerico sendo formada por 51 aminoácidos, e tem várias funções importantes, como regulação síntese de lipídios, da atividade enzimática e acima de tudo dos níveis de glicose no sangue e prevenção de hiperglicemia.

As estatísticas mostram que o número de casos de diabetes tipo **II** vem aumentando, em decorrência de fatores como aumento da expectativa de vida da população, mudança de hábitos, rotinas e estilo de vida característicos das grandes cidades, como consumo de alimentos industrializados e *fast-foods* por exemplo. Esses acabam sendo consumidos devido a praticidade desses alimentos e ao “frenesi cotidiano”. No entanto isso resulta em um fator de risco grave, pois a maioria desses alimentos são ricos em gorduras e carboidratos, que aliada ao sedentarismo, pode resultar em obesidade, e conseqüentemente, culminar na diabetes melito tipo **II** e outras doenças.

Foram realizados estudos em 200 países que revelaram dados alarmantes: no período de 1980 até 2014 houve aumento de 108 milhões para 422 milhões de casos, embora tenha havido aumento populacional de aproximadamente 40%, ou seja, o número de casos teve um acréscimo desproporcional (LANCET, 2016). Como consequência desse crescimento do números de casos, as despesas terapêuticas com pacientes elevaram-se para cerca de 825 bilhões de dólares anuais, já que muitos deles, além do próprio tratamento da diabetes, remediaram suas comorbidades como hipertensão, AVC, dislipidemia, dentre outras (LIPSKA, HIRSCH e RIDDLE, 2017).

As injeções de insulina após as refeições ainda constituem o principal tratamento para diabetes, no entanto, as aplicações contínuas de insulina podem acarretar em ganho de peso e favorecer a resistência à insulina no longo prazo (FINEMAN, 2005). Por isso, as doses deveriam ser restritas ao mínimo necessário, pois a insulino terapia leva progressivamente ao aumento das doses, o que, diminui a qualidade de vida dos pacientes e o desinteresse pela terapia.

A IAPP (Islet Amyloid Polypeptide), ou simplesmente amilina, é uma pequena proteína constituída por 37 aminoácidos, ou resíduos. Ela é secretada pelas células  $\beta$  do pâncreas (WESTERMARK, ANDERSON e WESTERMARK T., 2011), juntamente com sua parceira sinérgica, a insulina. A amilina possui várias funções no organismo, como regular a saciedade e o esvaziamento gástrico, sendo um fator importante para manter a homeostase energética. Ela foi descoberta durante procedimentos de autópsia em pacientes mortos com a doença, onde foi constatado formação de aglomerados amilóides no pâncreas endócrino de humanos diabéticos tipo II. Distúrbio esse observado em mamíferos suscetíveis, como os macacos (LUCA *et al.*, 2007).

A notável função metabólica da amilina, aliada à sua característica amiloidogênica, ou seja que forma amiloides ou fibrilas, incentivou pesquisas por análogos mais estáveis, menos propensos à agregação e que aumentassem a solubilização com insulina. Uma coformulação em dose única com insulina seria terapeuticamente preferível, já que reduziria o número de aplicações diárias. Atualmente existem análogos já aprovados pelo *Food and Drug Administration* (FDA) como a Pramlintide ou sIAPP (WESTERMARK, ANDERSON e WESTERMARK T., 2011), que é constituída pela substituição de duas Serinas e uma Alanina por Prolinas em posições estratégicas da hIAPP, respectivamente (i.e., S28P, S29P e A25P) com intuito de aumentar a sua resistência à agregação.

Em nosso estudo simulamos a co-formulação de insulina com amilina humana, e alguns de seus análogos, para avaliar a estabilidade e eventual efetividade dessas formulações. Para tal, utilizamos o *software* “*Simple Molecular Mechanics for Proteins*” (SMMP) no supercomputador Santos Dummont (LNCC). Em seguida esses dados computacionais foram estudados via análises microcanônicas, bem como estruturais. Em nossa proposta, enfocamos no análogo sIAPP<sup>+</sup>, descoberto computacionalmente



em (FRIGORI, 2017), que apresenta uma solubilidade aprimorada graças a uma substituição de uma Serina por Arginina na posição 20 da Pramlintida. Dado que essa Arginina é um resíduo positivamente carregado, e estando estrategicamente posicionado, este mutante intitulado S20R Pramlintide foi denominado como sIAPP<sup>+</sup>.

Posteriormente utilizamos o pacote PHAST para extrair e analisar as propriedades termodinâmicas. Por meio da mecânica estatística microcanônica, estudamos o aparecimento de heterodímeros eventualmente formados nas transições de fase de agregação de Insulina com mutantes de Amilina. Em seguida calculamos as seguintes grandezas termodinâmicas: Energia livre de Helmholtz ( $\Delta F$ ), o calor latente de agregação ( $L$ ) e temperatura de transição ( $T_c$ ). Tais grandezas foram utilizados nas análises e caracterização fenomenológica dessas co-formulações.

Também confrontamos o perfil de solubilidade das coformulações de insulina com análogos da amilina, com o potencial eletrostático molecular calculado a partir da equação de Poisson-Boltzmann, o que propiciou adicionalmente verificarmos termodinamicamente a estabilidade dessas soluções. Após criteriosa análise dos dados, propomos e discutimos uma possível alternativa para a co-formulação da insulina com o análogo da amilina mais físico-quimicamente estável. Para completar o estudo calculamos os tempos característicos de agregação, denominados “*time lags*”, baseados em uma analogia com a teoria clássica de nucleação (CNT)( FOKIN, YURISTSYN e ZANOTTO, 2005).

## 1.2 OBJETIVO GERAL

Simular computacionalmente com métodos microcanônicos, ou seja, via simulações de Monte Carlo multicanônicas (MUCA) (BACHMANN, 2014) o processo físico-químico de formação de heterodímeros entre Insulina humana e mutantes biotecnológicos da Amilina (sIAPP e sIAPP<sup>+</sup>).

## 1.3 OBJETIVOS ESPECIFICOS

- Simular computacionalmente com MMC via multicanônicas (MUCA) (BACHMANN, 2014);
- Analisar a formação de heterodímeros entre Insulina humana e mutantes biotecnológicos da Amilina (sIAPP e sIAPP<sup>+</sup>);
- Analisar e comparar as características estruturais e propriedades físico químicas das coformulações;
- Analisar a amiloidogenicidade intrínseca e a estabilidade contra eventuais agregações nas diversas co-formulações;
- Finalmente, confrontar os tempos de agregação (*times lags*) das co-formulações de Insulina com os respectivos análogos da Amilina para aferir seu potencial no desenvolvimento de futuras propostas terapêuticas.

## 2 PROTEÍNAS: UMA BREVE REVISÃO BIBLIOGRÁFICA

Existe um grande interesse em estudar os mecanismos funcionais das interações de insulina com amilina. Nesse sentido, já houve esforços por exemplo para investigar a inibição da agregação de amilina por meio de moléculas orgânicas (NEDUMPULLY-GOVINDAN, *et al.* 2016).

Recentemente foram realizadas análises físico-químicas de Pramlintide *in vitro* com pH análogo ao fisiológico, em solução tampão, em que foram evidenciados sinais de agregação por indícios de precipitação em testes de fluorescência com tioflavina (ThT), que revelam fases sólidas na solução. Ou seja, o experimento demonstrou que a solução de Pramlintide não seria potencialmente apta para uma eficaz co-formulação com insulina, pois pode levar a formação de agregados e precipitados (CABRAL e T.R LIMA, 2017) de modo mais brando mais ainda assim similar à hIAPP em (WANG, *et al* 2014).

Nessa revisão serão apresentados os temas que compõe o referencial teórico que fundamenta a pesquisa, abrangendo em seu conteúdo, o estudo das proteínas, aminoácidos, peptídeos, bem como, suas interações químicas moleculares, estruturações e suas propriedades termodinâmicas.

### 2.1 ESTUDO DOS AMINOÁCIDOS, PEPTÍDIOS E PROTEÍNAS

Cadeias de aminoácidos podem condensar-se formando oligopeptídeos e polipeptídeos. Se duas cadeias polipeptídicas são idênticas, a proteína é chamada de oligomérica. Algumas proteínas são caracterizadas por grupos de unidades polipeptídicas de cadeias  $\alpha$  e  $\beta$  idênticas. Assim, cada unidade  $\alpha$  é pareada de modo idêntico com uma unidade  $\beta$  dentro da estrutura da proteína, onde pode ser considerada tanto um tetrâmero de quatro unidades de polipeptídeos, quanto um dímero de protômeros  $\alpha\beta$  (LEHNINGER, 2014, pg 87-88).

Nas proteínas naturais são encontrados vinte diferentes tipos de aminoácidos. Eles diferem-se uns dos outros em suas cadeias laterais ou grupos R, em estrutura, tamanho, carga elétrica e estrutura molecular, como a prolina por exemplo que possui estrutura cíclica específica. Tais características também influenciam as propriedades das proteínas, como a solubilidade molecular, por exemplo, podem ser usadas nos compostos para facilitar a quebra de ligações, propiciando um aumento na dissolução dos compostos em solução.

A solubilidade de compostos proteicos em meio aquoso depende da ocorrência de grupos polares na cadeia lateral, como, nos grupos hidroxílicos: serina e treonina. Em contrapartida as cadeias laterais hidrofóbicas são insolúveis em um ambiente polar, situações típicas ocorrem em aminoácidos que possuem anéis aromáticos como: fenilalanina e triptofano. Outros, são alifáticos como alanina e leucina, isto é, contêm átomos de hidrogênio e carbono em suas cadeias laterais.

Eletrostaticamente os aminoácidos que possuem carga, contribuem para definir a carga total de uma proteína em um meio aquoso com solução neutra (BACHMANN, 2014). Teoricamente é de suma importância apontar essa condição, tendo em vista que em nosso estudo, o análogo da amilina proposto tem uma serina substituída por arginina carregada positivamente, a fim de melhorar sua solubilidade.

## 2.2 CARACTERIZAÇÃO QUÍMICA DOS AMINOÁCIDOS

Os aminoácidos podem ser classificados em três classes de acordo com as características químicas de suas cadeias laterais como mostrados no **Quadro 1**.

- **Aminoácidos hidrofóbicos ou apolares:** que são repelidos por serem insolúveis em água.
- **Aminoácidos hidrofílicos ou polares:** são atraídas ou dissolvidas pela água.
- **Aminoácidos anfipáticos:** suas moléculas apresentam, ao mesmo tempo, características polares e apolares.

**Quadro 1: Classificação dos aminoácidos, características químicas.**

<b>HIDROFÓBICOS</b>	<b>HIDROFÍLICOS</b>	<b>ANFIPÁTICOS</b>
Valina	Aspariginina	Glicina
Leucina	Ác.glutâmico	Alanina
Isoleucina	Glutamina	Serina
Metionina	Histidina	Treonina
Fenilalanina	Lisina	Tirosina
	Arginina	Triptofano
	Ác. aspártico	Cisteína
		Prolina

Fonte: Elaborada pelo autor.

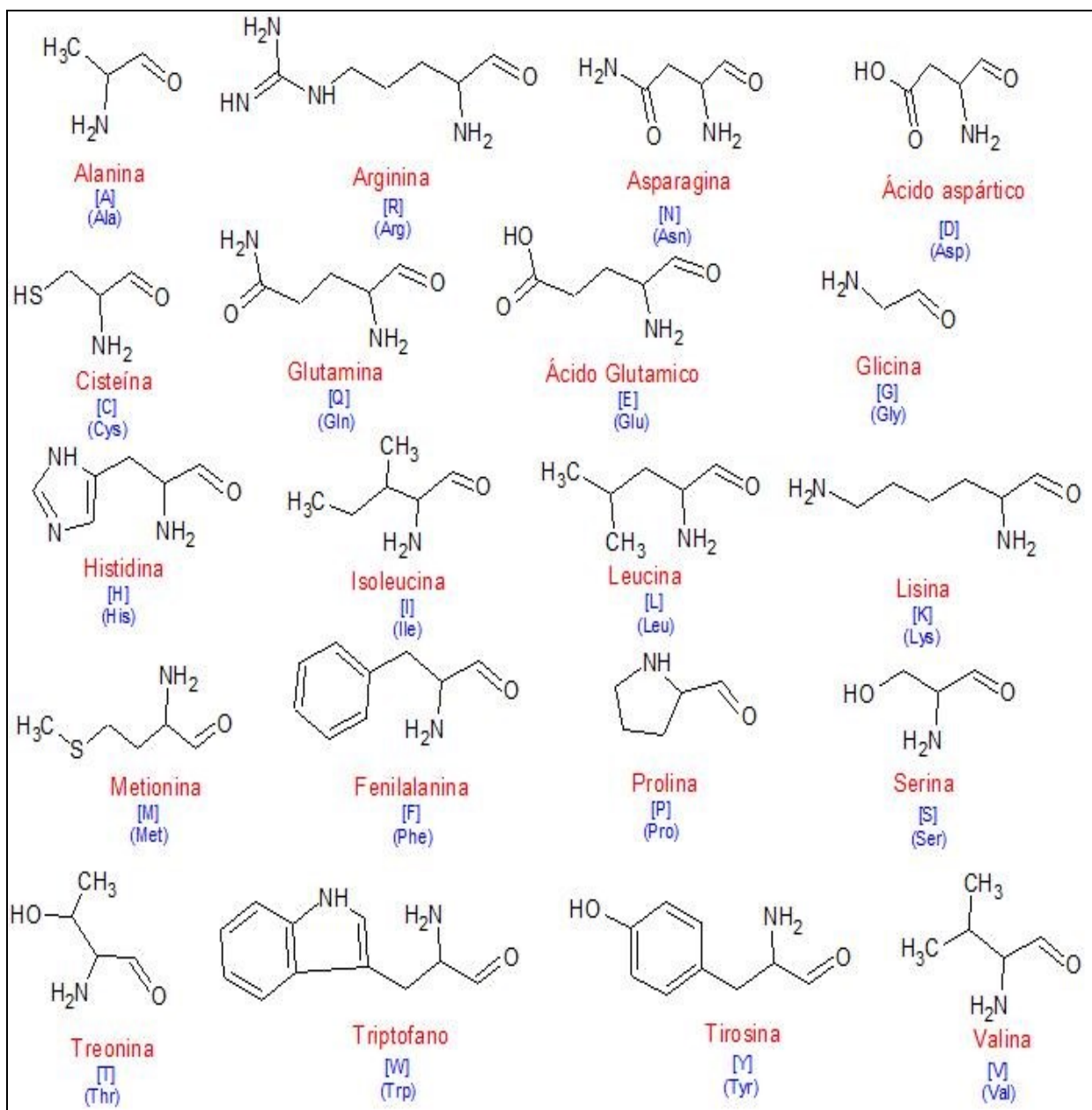
As proteínas são as moléculas orgânicas mais abundantes nas células e correspondem a cerca de 50% ou mais de seu peso seco. São encontradas em todas as partes de todas as células, tendo funções fundamentais em seu desenvolvimento. Em virtude desta importância qualitativa e quantitativa, as proteínas têm sido largamente estudadas (HUANG, 2005) (LANDAU e BINDER, 2009) (LEHNINGER, 2014) e seus segredos desvendados, no que diz respeito à sua síntese ou aproveitamento metabólico.

As proteínas são sequências encadeadas de aminoácidos, ou seja, ligados entre si. A quantidade de aminoácidos, sua repetição e conformação espacial caracterizam uma proteína e sua função específica, onde as mesmas são constituídas e compostas por moléculas de carbono, hidrogênio, oxigênio e nitrogênio. São ainda de origem animal, vegetal e sintética, sendo essenciais para manutenção dos organismos.

Contudo é necessário conhecer as peculiaridades das ligações químicas que levam à formação de sistemas macromoleculares microscópicos, provocando a estruturação de inúmeros complexos moleculares. Sendo que para tentar compreender esses processos, é de bom alvitre empregar – se ferramentas estatísticas para analisar a formação dos compostos moleculares como às proteínas. Podendo assim, desvendar o que provoca o surgimento de doenças e sua relação com as proteinopatias, para que

enfim, possa-se pensar em soluções potenciais para as disfunções orgânicas como relataremos em nosso estudo. A **Figura 1** mostra as fórmulas moleculares e abreviações dos 20 aminoácidos naturais.

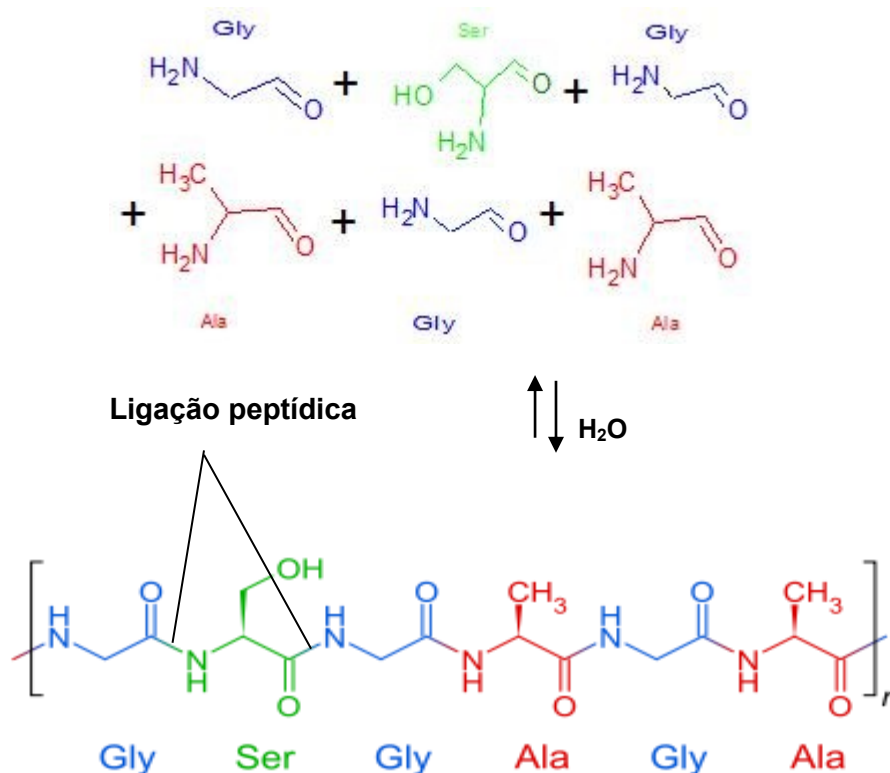
**Figura 1: Fórmula molecular e abreviações dos 20 aminoácidos.**



Fonte: Elaborada pelo autor, utilizando o software *ChemSketch* da ACD/Labs, 2019.

Os peptídeos, por sua vez, são constituídos a partir de ligações de aminoácidos por compartilhamento covalente de elétrons, dando origem a ligações peptídicas. A ligação de duas moléculas peptídicas resultará em um dipeptídeo liberando uma molécula de água conforme esquematizado na **Figura 2**: Representa a fibroína, formada pelos aminoácidos respectivos, constitui 70% de uma proteína polimerizada por (*n*) repetições secretada pelo bicho da seda ( *Bombyx mori*), que é amplamente utilizada para produção de uma das fibras têxteis mais nobre: a seda. Os outros 30 % constitui a sericina, utilizada para fabricação de cosméticos e fins medicinais.

**Figura 2: Ligação covalente do dipeptídeo ( Fibroína).**



Fonte: Elaborada pelo autor, utilizando o software *ChemSketch* da ACD/Labs, 2019.

## 2.3 O ENOVELAMENTO OU DOBRAMENTO DE PROTEÍNAS

A conformação de uma proteína é o arranjo espacial dos átomos que constituem uma distribuição tridimensional. Vários fatores contribuem para a existência de diferentes conformações, entretanto, a natureza sempre tende a assumir conformações termodinamicamente mais estáveis.

Do ponto de vista físico, isto significa dizer que a energia livre de Gibbs (G) da proteína deve ser mínima. Deste modo, define-se a estrutura nativa como sendo a conformação enovelada de uma proteína em seu estado de mínima energia livre e que desempenha alguma função biológica ativa. A conformação nativa da proteína é o estado mais estável de configuração espacial, onde pequenas mudanças no ambiente que a cerca podem causar alterações estruturais e afetar sua função biológica (ALMEIDA, 2016).

O enovelamento é um processo bioquímico de extrema importância para a formação das proteínas, pois, é através dele que a proteína assume sua estrutura específica tridimensional nativa e funcional. Ao processo contrário denomina-se desnaturação, nesse caso, a proteína é forçada a perder sua forma nativa, tornando-se uma cadeia de aminoácidos sem funcionalidade e insolúvel (LEHNINGER, 2014).

ANFINSSEN (1973) denominou “Hipótese Termodinâmica” o processo de dobramento da proteína conhecido como *foldig*, porque ocorreria de forma aleatória em tentativas e erros até testar todas as posições possíveis de suas ligações e encontrar seu estado nativo, ou seja, termodinamicamente mais favorável. Isso só é possível devido ao trabalho de proteínas auxiliares: as chaperonas que redobram as proteínas, quando as mesmas não concluíram o processo, ou que concluíram de maneira inadequada.

LEVINTHAL (1968) levantou a seguinte questão: que o processo de *foldig* para encontrar a estrutura nativa das proteínas é tão complexo que dependendo da quantidade de aminoácidos envolvidos nos processos iterativos e de sua aleatoriedade, pode levar a milhares de anos ou em casos extremos, inclusive superior a idade do universo. Esse fato logo ficou conhecido como seu paradoxo, que somente foi explicado



anos depois com a “Teoria do Funil” energético que prediz um perfil termodinâmico propício sendo viesado na direção da estrutura nativa.

No entanto, quando ocorrem esses estados não dobrados formam agregados que se depositam nas células, podendo resultar desse processo o desencadeamento de doenças degenerativas, conhecidas como proteinopatias, ou seja, Diabetes, Parkinson, Alzheimer dentre outras (LEHNINGER, 2014 pg 143).

“O desdobramento de proteínas seguido por agregação amilóide é uma fonte adicional para danificar as redes celulares. No caso da doença neurodegenerativa de Alzheimer (DA), que causa demência, é motivada quando experimenta mudanças estruturais e tende a formar agregados peptídicos de beta-amilóides ( $A\beta$ ) em alta concentração (BACHMANN 2014, pg.01).”

Durante o processo de enovelamento, as proteínas recém sintetizadas nos ribossomos passam por *foldings* estruturais, dirigidos pelo conhecido “colapso hidrofóbico” visando atingir sua máxima estabilidade conformacional. Os estados enovelados são termodinamicamente separados dos não enovelados por uma barreira de energia de ativação ( $\Delta G$ ), ou seja, quanto maior a barreira de ativação mais lenta é a cinética de reação e o processo de conformação molecular. Nessa etapa podem surgir fenômenos cooperativos de mudanças de fases, sendo que esse fenômeno é importante objeto de investigação do comportamento físico-químico das proteínas e suas finalidades.

A capacidade de uma molécula “grande” consistindo de centenas a milhares de átomos para formar estruturas estáveis espontaneamente é normalmente chamado de “cooperatividade”. Esse termo não é bem definido e poderia ser facilmente substituído por “emergência” ou “sinérgica” - noções que foram cunhados em outros campos de pesquisa para a mesma característica misteriosa de efeitos de ordenação (BACHMANN, 2014 pg.14).

O mecanismo de *folding*, além de envolver aspectos estruturais, também abrange particularidades energéticas tais como energia potencial da proteína, números de ligações de hidrogênio, raio de giração, energia de solvatação, bem como, o volume do solvente (ALMEIDA, 2016). Algumas dessas propriedades serão objeto de análise de nosso estudo posteriormente.

## 2.4 FUNCIONALIDADES DAS PROTEÍNAS

As proteínas apresentam uma extensa lista de funções biológicas e são encontradas nas estruturas celulares, substâncias intersticiais, anticorpos, entre outros. Entre as funções que podem ser atribuídas às proteínas, destacam-se:

- **Transportadoras:** Transporte de oxigênio para o corpo (hemoglobina);
- **Anticorpos:** Atuam na proteção imunológica contra organismos patogênicos;
- **Catalizadora:** Para acelerar as reações químicas como no caso das enzimas;
- **Contração muscular:** actina e miosina;
- **Estrutural:** Confere rigidez, consistência e elasticidade aos tecidos;
- **Energética:** Fornecimento de energia pelos canais que formam as proteínas;
- **Hormonal:** Crescimento e regulação metabólica.

### 2.4.1 CLASSIFICAÇÃO DAS PROTEÍNAS

Podem ser classificadas como:

- **Proteínas simples:** Apresentam apenas aminoácidos em sua formação;
- **Proteínas conjugada:** Além de aminoácidos, apresentam um radical de origem não peptídica, que é denominado de grupo prostético;
- **Proteínas derivadas:** Não são encontradas na natureza e são conseguidas graças a processos de degradação de proteínas simples ou conjugadas.

#### 2.4.1.2 Classificação Morfológica das Proteínas

A forma das proteínas, suas características geométricas, as seções microscópicas longitudinal e transversal, também podem influenciar nas suas funcionalidades e são classificadas morfológicamente como:

- **Globulares:** São as mais comuns, sensíveis a desnaturação e possuem funções metabólicas como a hemoglobina;
- **Fibrilares:** Possuem característica de serem filamentosas, por isso tendem a agregarem com mais facilidade como a queratina nos animais;
- **Membranosas:** Função de adesão celular, e transporte como as glicoproteínas;
- **Proteínas Intrinsecamente Desestruturadas (IDPs):** são proteínas que são caracterizadas por sua falta de estrutura terciária estável, quando na forma de uma cadeia peptídica, isolada sob condições fisiológicas *in vitro*. Sua descoberta desafiou o paradigma clássico da estrutura proteica, que considerava que uma estrutura bem definida era essencial para a proteína desempenhar sua função, bem como a estrutura que define essa função.

#### 2.4.2 REPRESENTAÇÕES COMPUTACIONAIS DAS PROTEÍNAS

As proteínas, devido às suas estruturas complexas, necessitam de uma representação computacional robusta. É preciso prover informações tais como os comprimento e ângulos de ligações entre os átomos e informações acerca das posições dos átomos. Utiliza-se, frequentemente, os formatos de arquivo FASTA e PDB para guardar esses tipos de dados.

O formato FASTA é o modo mais comum para representar sequências de nucleotídeos ou sequências de aminoácidos para serem usadas em vários *softwares* em simulações computacionais. Seus arquivos normalmente, configuram as sequências de aminoácidos e são iniciados com o símbolo ">" (ALMEIDA, 2016).

Já os arquivos PDBs, são formados por banco de dados e não documentos, por isso devem ser abertos em programas específicos. Eles podem ser arquivos como no *Protein Data Bank*, que podem ser utilizados para estudar as estruturas de proteínas e por isso, esses arquivos são amplamente utilizados em pesquisas.

### 2.4.3 HIERARQUIA ESTRUTURAL DAS PROTEÍNAS

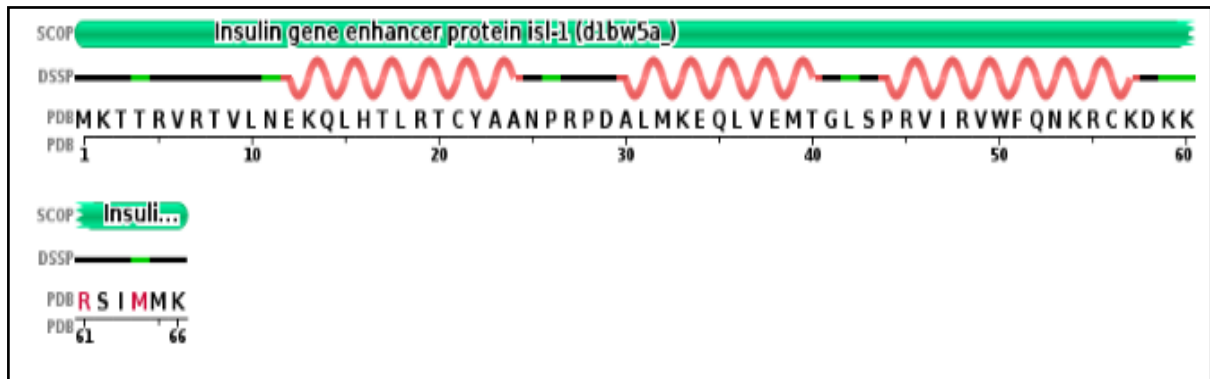
As proteínas que são constituídas por apenas uma cadeia são denominadas de monoméricas, no entanto, a maioria são formadas por muitos monômeros, ou seja, multimonoméricas. Vejamos como acontece o processo de constituição das proteínas bem como, as etapas de estruturação.

#### 2.4.3.1 Estrutura Primária

É o nível estrutural mais simples, são específicas de cada proteína e determinadas geneticamente pela constituição e sequência de aminoácidos. Sendo que quando estiverem com os aminoácidos com suas estruturas semelhantes denomina-se homólogos: Que são proteínas que derivam de um "ancestral comum". Podem estar presentes em uma mesma espécie, sendo derivado por duplicação de genes no genoma de um organismo, ou em espécies diferentes (OPEN STAX, 2018).

A **Figura 3** mostra um encadeamento encontrado no banco de dados *Protein Data Bank*, onde a estrutura primária da insulina de uma espécie de rato, o *Rattus Norvegicus* com 66 aminoácidos. O segmento de aminoácidos da estrutura primária contém informações que determinam por exemplo a formação das hélices alfa em (espiral vermelho) em intervalos de aminoácidos, que posteriormente caracterizarão, como ela assumirá as suas estruturas secundárias e terciárias (i.e. espaciais), via enovelamento da proteína e funções orgânicas específicas.

Figura 3: Exemplo de sequência de aminoácidos com 66 resíduos do PDB 1BW5.



Fonte: Protein Data Bank, 2019.

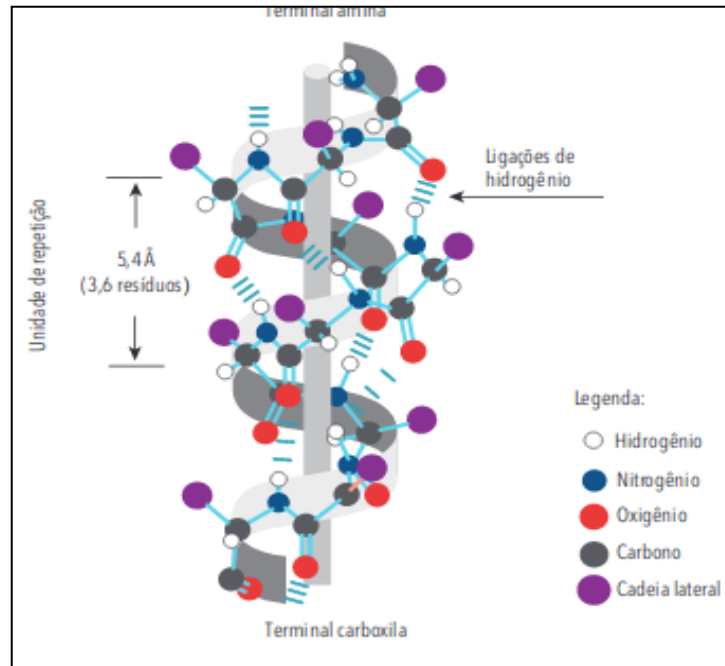
#### 2.4.3.2 Estrutura Secundária

Constituída entre o arranjo espacial do encadeamento de aminoácidos e a sequência primária dos aminoácidos. O arranjo secundário dos polipeptídios pode ocorrer quando os ângulos das ligações entre carbonos e seus ligantes se repetem ao longo da sequência da molécula sendo chamados de domínios.

Os tipos mais comuns de estruturas secundárias são:

**a) Hélices alfa:** Apresenta - se de forma enrolada como um espiral e os grupos R se fixam na parte externa da hélice alfa, onde estão livres para facilitar as interações. Conforme a **Figura 4**, as estruturas helicoidais da hélice alfa denominada de esqueleto de (*backbone*), constituída de átomos de hidrogênio, nitrogênio, oxigênio, carbono e uma cadeia lateral. Verificamos também que, para cada volta, percebe-se 3,6 resíduos e um espaçamento entre átomos na ordem de 5,4 Å.

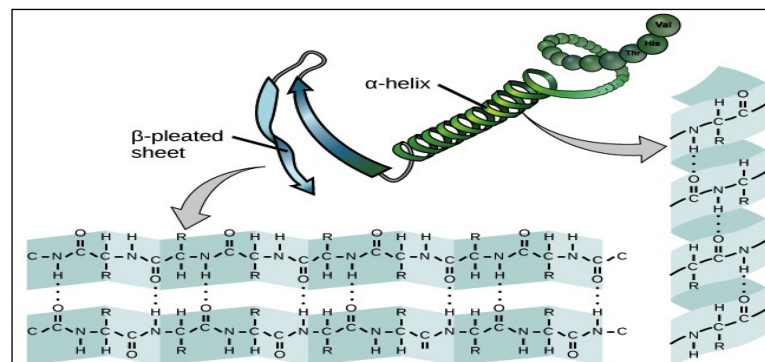
Figura 4: Estrutura (*backbone*).



Fonte: Adaptado de Almeida, 2016.

**b) Folhas betas:** Possuem uma formação similar a uma fita, formando uma estrutura parecida com uma folha dobrada. A prolina por exemplo, possui sua fórmula estrutural cíclica que não é compatível com formações em folhas beta. A **Figura 5** exemplifica a formação e união desses componentes das proteínas, com as ligações químicas, da carbonila, do grupo amina e da união com pontes de hidrogênio (OPEN STAX BIOLOGY, 2018).

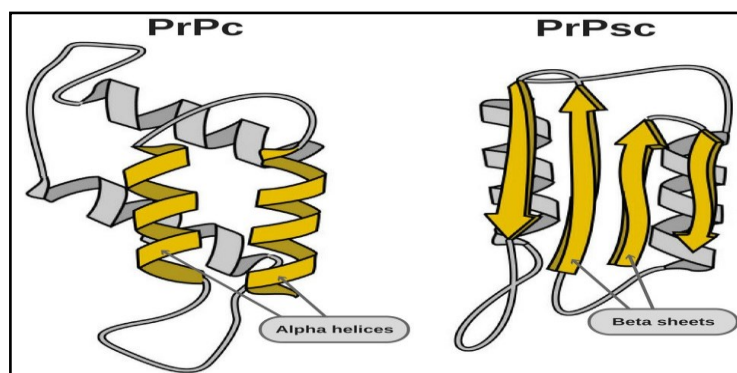
Figura 5: Folhas betas e hélices alfa com ligações químicas



Fonte: adaptado de Open Stax Biology, 2018.

Em seu enovelamento, as proteínas produzem uma conformação estrutural característica e verifica-se que, a cada grupo de aminoácidos, em uma respectiva repetição, corresponde uma conformação estrutural chamada de domínio. Os domínios usuais aparecem como estruturas secundárias típicas: hélices-alfa e folhas-beta conforme são vistas abaixo na **Figura 6**.

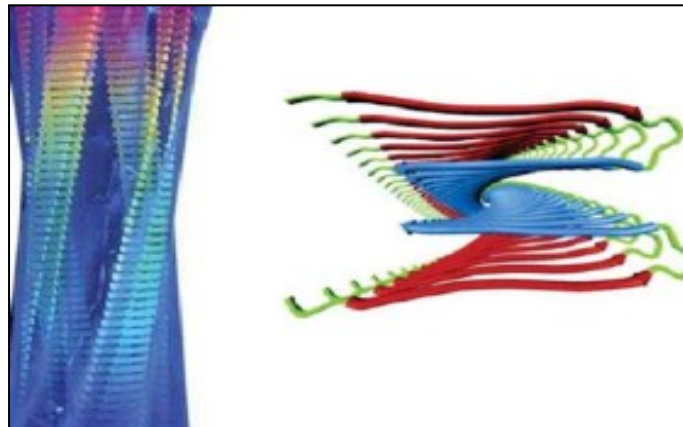
**Figura 6: Mecânismo priônico.**



Fonte: Science ABC, 2019.

A **Figura 6**, denominada de Príon, é um agente infeccioso composto por proteínas na forma aberrante. Inicialmente descoberta por Stanley Prusiner, e que está relacionada ao desenvolvimento de encefalopatias (SCIENCE ABC, 2019) por um mecanismo de inter-conversão patogênica (amilóide) de estruturas hélices-alfa em folhas-beta, em amarelo, ou seja, uma estrutura imita e replica a morfologia da outra.

As folhas betas microscopicamente podem apresentar-se morfologicamente em formação de fibrilas, que ao se encadearem transformam-se em deposições amiloides e que estão relacionados ao surgimento de doenças neurodegenerativas. A **Figura 7** apresenta respectivamente as vistas longitudinais e transversais das fibrilas.

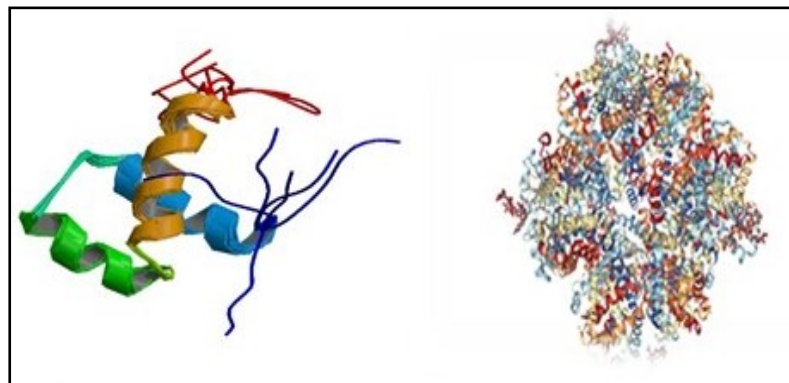
**Figura 7: Formação de fibrilas**

Fonte: Fonte: Adaptado The Lancet

#### 2.4.3.3 Estrutura Terciária e quaternária

A estrutura terciária pode ser entendida como um agrupamento de domínios, e ocorrem quando as estruturas primárias se dobram e conectam-se através de ligações e interações químicas como veremos mais adiante, formando ramificações responsáveis pela estabilização molecular das proteínas. Ao passo que as estruturas quaternárias ocorrem com a união de várias estruturas terciárias bem definidas, como é o caso da insulina PDB 1BW5 ou junção de duas ou mais subunidades alfa-beta como exemplo da hemoglobina 3WCU representadas na **Figura 8** respectivamente.

**Figura 8: Domínio terciária de um tipo de insulina, PDB 1BW5 e Estrutura hemoglobina 3WCU respectivamente.**



Fonte: Extraídos do [Protein Data Bank](https://www.rcsb.org/).



## 2.5 PROPRIEDADES FÍSICO-QUÍMICAS DAS PROTEÍNAS E SUAS INTERAÇÕES MOLECULARES

O estudo dos mecanismos das interações moleculares é de extrema importância, devido ao seu controle e sincronização na síntese e no metabolismo dos carboidratos, das gorduras e proteínas ao longo do tempo. Os distúrbios nas interações hormonais, podem levar os organismos a tornarem-se propensos a desenvolverem doenças (LEHNINGER, 2014). Para tanto as células precisam sintetizar simultaneamente milhares de moléculas nas exatas proporções requeridas para manter o equilíbrio orgânico, como por exemplo, a homeostase da glicose.

Quando os aminoácidos são submetidos a um meio aquoso, tornam-se termodinamicamente propícios a rearranjarem-se e adquirem propriedades físico-químicas específicas. Os complexos microscópicos de proteínas visam alcançar maior estabilidade para formação das biomoléculas com estruturas geométricas específicas implicando em elevação da entropia do solvente. No entanto, esses elementos, podem sofrer restrição estruturais durante tais transições de fase (BACHMANN, 2014). Assim, as proteínas precisam de um contrabalanço energético para estabilizar essa elevação de entropia. Nesse caso, a solução pode se dar através de determinadas interações, como as eletrostáticas, as de Van der Waals, hidrofóbicas e as iônicas. As interações mais fracas e de longo alcance, são muito importantes para o enovelamento das proteínas, já que tais ligações podem ser rompidas deixando a estrutura resultante com menor energia (LEHNINGER 2014 pg 116).

Muitos tipos de interações químicas podem ocorrer dentro de uma molécula proteica para garantir a estabilidade das cadeias polipeptídicas. As mais fortes são as ligações covalentes, como a que ocorre entre dois aminoácidos de cisteína, que se unem covalentemente através de ligações de dissulfetos.

Observemos alguns tipos de interações intermoleculares e intramoleculares:

- **Ligações Covalentes:** As ligações covalentes são baseadas em pares de elétrons compartilhados para estabilizar a composição química da molécula.
- **Ligações não-covalentes:** As interações não-covalentes são induzidas por forças de Van der Waals, eletrostáticas potenciais entre cargas parciais de

átomos, potenciais de torção, arranjos dipolo-dipolo, formação de ligações de hidrogênio e forças intermoleculares efetivas, como, por exemplo, causada pelo efeito hidrofóbico, devido à interação da proteína com o solvente (BACHMANN pg 10, 11).

- **Ligações de Hidrogênio:** Grande número de ligações de hidrogênio são formadas no interior e na superfície das proteínas. Além de formar ligações de hidrogênio entre si, os grupos polares das cadeias laterais dos aminoácidos podem interagir com a água ou com o esqueleto polipeptídico contribuindo, moderadamente para direcionar as ligações eletrostáticas entre moléculas polares, ou seja, com potencial:

$$P = \left( \frac{1}{R^{12}} - \frac{1}{R^{10}} \right),$$

que acontece quando um grupo doador possui um átomo de hidrogênio que se liga a um átomo de alta eletronegatividade podendo ser, intramolecular ou intermolecular, além de contribuir decisivamente para as configurações de folhas  $\beta$ . Onde R é a distância entre os átomos de hidrogênio.

- **Interações Hidrofóbicas:** São as forças não covalentes mais importantes para a estabilidade da estrutura enovelada. Ocorrem entre aminoácidos de cadeia lateral hidrofóbica, excluindo e afastando as moléculas de água no momento das ligações. Essas interações são formadas por substâncias apolares e possuem tendência de se agregarem quando imersas em soluções aquosas, repelindo a água (ALMEIDA, 2016).
- **Interações eletrostáticas (ligações iônicas):** Ocorrem entre aminoácidos que possuem carga positiva ou negativa na cadeia lateral, ou seja, Potencial Coulombiano ou seja:

$$P = \frac{1}{R}.$$

Onde R é a distância entre os átomos.

- **Forças de van der Waals:** São forças de atração inespecífica que ocorrem quando dois átomos quaisquer estão próximos. Apesar dessas forças serem comparativamente fracas em relação as demais, o efeito cumulativo de

numerosas interações tem substancial influência para a estabilidade da estrutura enovelada. Apresentam forças de repulsão para pequenas distâncias e de atração para grandes distâncias. A natureza desse fenômeno pode ser descrita via potencial de Lennard-Jones, ou seja,

$$P = \left( \frac{1}{R^{12}} - \frac{1}{R^6} \right)$$

Onde R é a distância entre os átomos (ALMEIDA, 2016).

- **Ligações salinas:** São interações iônicas que ocorrem nas cadeias laterais dos aminoácidos com intuito de obter estabilidade;
- **Ligações de dissulfeto:** Ocorre interações entre S-S (enxofre-enxofre) e se ligam entre ramificações. Em alguns tipos de insulina por exemplo as ramificações ocorrem entre as cadeias de aminoácidos também com a intenção de alcançar estabilidade.

Nesse contexto, podemos ressaltar que o estudo *in silico* de mecanismos moleculares, associados à origem de distúrbios e/ou doenças, vêm se tornando de fundamental importância no entendimento, controle e cura de doenças ocasionadas por proteínas (proteinopatias). E em nosso estudo, utilizaremos a termoestatística microcanônica como metodologia que permite inferir o comportamento termodinâmico das proteínas a partir dessas interações moleculares.

### 3 PROPRIEDADES DA TERMODINÂMICA DE EQUILIBRIO

Entende-se o conceito termodinâmico de entropia: como a razão entre a variação da quantidade de calor a uma dada temperatura. A entropia mede o grau de desordem molecular de um sistema, e está relacionada ao seu número de configurações ou microestados, ou seja, de quantas maneiras as partículas (átomos, íons ou moléculas) podem se distribuir incluindo configurações translacionais, vibracionais, rotacionais e eletrônicas. Entropia também é geralmente associada à aleatoriedade, dispersão de matéria e energia, de um sistema termodinâmico. Fisicamente é a função de estado que rege a segunda lei da termodinâmica.

Dimensionada em (J/K), estabelece seu aumento para processos espontâneos em sistemas isolados. Para sistemas abertos, deve-se estabelecer que a entropia do universo (sistema e suas vizinhanças), deve aumentar até atingir um valor máximo no estado de equilíbrio. A segunda lei da termodinâmica foi concebida com o estudo de processos cíclicos que convertiam calor em trabalho, que surgiram nos primórdios da revolução industrial (SMITH; NESS; ABBOTT, 2007, pg 29-50).

No entanto, percebeu-se que o calor não poderia ser completamente convertido em trabalho em máquinas térmicas, havendo a necessidade de perda mínima de calor para que o processo voltasse ao estado original. Assim, a variação de entropia de um processo mede a parcela de energia que não pode mais ser convertida em trabalho em transformações termodinâmicas a dada temperatura.

Por esse motivo, a mensuração da variação de entropia está diretamente relacionada a processos reversíveis, uma vez que eles estão associados com a máxima energia que um sistema pode perder realizando trabalho, ou seja, chamado de trabalho máximo ou útil. Portanto a variação de entropia mede o grau de irreversibilidade de um processo.

O fato é que, conhecendo a entropia, é possível relacionar e mensurar outras grandezas, como a energia interna ( $\epsilon$ ), a quantidade de partículas (N) e a natureza dessas partículas que integram o sistema, bem como, o seu volume (V), calor latente

de transição ( $\Delta L$ ) e certamente variações de entropia ( $S$ ) (SMITH; NESS, ABBOTT; 2007, pg 29-50).

Isto é, a entropia  $S(\epsilon, N, V)$ , quando expressa em função da energia interna ( $\epsilon$ ) do sistema, da quantidade de partículas ( $N$ ), e do volume ( $V$ ), é uma função de estado termodinâmica fundamental para um sistema térmico, da qual pode-se, mediante a transformada de Legendre, obter-se outras equações fundamentais como:

- A Entalpia  $H(S, P, N)$ ;
- Energia Livre de Helmholtz  $F(T, V, N)$ ;
- Energia Livre de Gibbs  $G(T, P, N)$ ;
- Grande Potencial Canônico  $U(T, P, m)$ .

Em sistemas termodinâmicos mais complexos, a exemplo dos sistemas magnéticos, outras variáveis podem figurar na equação entrópica fundamental e nas outras equações fundamentais, entretanto os conceitos de equação fundamental e transformada de Legendre permanecem os mesmos. Assim, a entropia nos estados de equilíbrio termodinâmico, permite estabelecer quaisquer outras das demais equações de estados fundamentais.

Posteriormente por volta de 1876, o físico Willard Gibbs chegou à conclusão de que o conceito de energia disponível ( $\Delta G$ ) em um sistema termodinâmico pode ser matematicamente obtido através da subtração entre a energia perdida ( $T\Delta S$ ) e a variação da energia total do sistema: a entalpia ( $\Delta H$ ). Estes conceitos foram aprimorados posteriormente por James Clerk Maxwell 1871 e Max Planck 1903 (CALLEN 1985, pg 5-56). Atualmente estes conceitos são ferramentas de investigação dos fenômenos físico-químicos da natureza, inclusive das proteínas, pois é sabido que essas quantidades controlam e determinam os processos conformacionais das proteínas desde seu enovelamento até sua estruturação, funcionalidades, e relacionando-se também com seus distúrbios funcionais.

## 4 O PRINCÍPIO DA TERMOESTATÍSTICA MICROCANÔNICA

O estudo da mecânica estatística, ou termoestatística, foi desenvolvido para explicar as propriedades termodinâmicas de sistemas macroscópicos que são formados por componentes microscópicos. Por exemplo, podemos citar um recipiente com gás em expansão que é composto por diferentes moléculas. Assim, devido à complexidade da análise de cada componente microscópico, o problema passou a ser tratado de forma probabilística, por isso a denominação “estatística”.

Esta abordagem pode ser aplicada a conjuntos de átomos, moléculas ou proteínas dentre outros sistemas, tendo como precursor, o físico Austríaco Ludwig E. Boltzmann (Viena, 20/2/1844 - Duino, 5/9/1906) como seu proponente (NEWMAN e BARKEMA, 1999). Na abordagem microcanônica, de Boltzmann, o “Princípio Ergódico” permite-nos substituir médias termodinâmicas por médias temporais, como explicado por Newman:

Para um sistema clássico que tem um continuum de estados de energia, pode haver um conjunto de estados degenerados através dos quais o sistema passa, e uma média sobre esses estados às vezes pode dar uma boa resposta para certas propriedades do sistema. Tais conjuntos de estados degenerados formam um conjunto microcanônico. O caso mais geral nós consideramos aqui, em que há um reservatório térmico causando a energia do sistema flutuar, é conhecido como um conjunto canônico (NEWMAN e BARKEMA 1999, pg 02).

Nesse contexto, é imprescindível a função de estado termodinâmica denominada entropia, haja visto que, em nível microscópico, ela é associada à idéia de desorganização estatística de sistemas físicos isolados, onde evidencia-se a relação direta do macro para o micro e equivalência com essas grandezas extensivas. Ou seja, em um sistema reversível, a entropia  $S(Q,T)$  é a função de estado cuja variação infinitesimal é igual à razão entre o calor infinitesimal trocado com meio externo ( $Q$ ) e a temperatura absoluta do sistema ( $T$ ). O equilíbrio termodinâmico, nesse contexto, é realizado pela maximização da entropia do sistema (FRIGORI, 2010).

## 4.1 O ENSEMBLE CANÔNICO

Considerando um reservatório (R) fechado em um banho térmico, com temperatura fixa (T) e um sistema de interesse (s) acoplado ao reservatório em que o calor é transferido por condução, com mesmos volumes de partículas. Este tipo de sistema é chamado de ensemble canônico. Nesse sistema os microestados ocorrem com probabilidade  $P_j$ , sendo “j” o grau de ocupação do microestado, em que a energia do sistema quando está no microestado  $j$  é denominada  $\epsilon_j$ . Geralmente, esses microestados podem ser considerados como estados quânticos discretos do sistema e suas probabilidades normalizadas são expressas como:

$$P_j = \exp\left(\frac{-\epsilon_j}{k_B T}\right) / Z$$

em que o fator de normalização (Z) é denominado como “função de partição” canônica. Quando o complexo (R+s) estiver em equilíbrio as probabilidades de um microestado “j” de R e s terem as mesmas propriedades são iguais (SALINAS, 2005).

Ocorre que, em algum momento, o banho como um todo assumirá uma temperatura final de equilíbrio. Isso não significa que, em equilíbrio, a energia não será mais alterada, pois através do acoplamento do banho térmico, o sistema pode ganhar energia do banho, por flutuações e perder energia por dissipação.

Assim, a energia de um microestado, atual do sistema mudará pela interação com o banho térmico, formando um novo microestado, que possui uma energia diferente. Considerando que as flutuações energéticas, podem, em princípio, ter algum valor, a energia interna do sistema ( $\mathcal{E}$ ) é constante em equilíbrio térmico em uma dada temperatura (T) constante, porque  $d\mathcal{E} = T.dS$  e a entropia (S) atinge seu valor máximo em equilíbrio. Por fim vale ressaltar que no equilíbrio, a Energia Livre de Helmholtz (F) conecta-se naturalmente à função de partição canônica (Z), como em (SALINAS, 2005) que veremos a seguir, pela relação fundamental:

$$F = -k_B T \ln(Z).$$

## 4.2 O ENSEMBLE MICROCANÔNICO

Considerando retirar uma amostra representativa do ensemble canônico com seus parâmetros termodinâmicos e somar ponderadamente, com o peso de Boltzmann todos os microestados ( $\Omega$ ), ou seja, definir todos os microestados chamados soma da partição isolados, isto é:

$$Z = \sum \Omega e^{-\beta \cdot \epsilon}$$

Sendo  $Z$  a função partição e  $\beta=1/k_B T$ . Termodinamicamente, o conjunto microcanônico corresponde a um grande sistema isolado, isto é, para um sistema quase sem flutuações energéticas, assumindo que sua energia total  $U \approx \epsilon \approx$  constante, então a probabilidade para realização das propriedades termodinamicas em todos microestados será identica (i.e. estados equiprovaveis) isto é, obtem-se conexão entre a probabilidade de ocorrência dos microestados ( $\Omega$ ) com a grandeza termodinâmica entropia ( $S$ ), através da famosa fórmula de Boltzmann (SALINAS, 2005):

$$S = k_B \log \Omega$$

## 4.3 TRANSIÇÕES DE FASES TERMODINÂMICA

O estudo das transições de fase tem sido desenvolvido por meio do ensemble Canônico e Microcanônico. A função termodinâmica que caracteriza transições de fase no conjunto canônico da mecânica estatística é a energia livre de Helmholtz ( $F$ ). Ela é obtida da energia interna ( $\epsilon$ ) e da entropia ( $S$ ), em um sistema à temperatura ( $T$ ), por meio de uma transformação de Legendre, ou seja:  $F = \epsilon - T \cdot S$ . Uma vez conhecida a energia livre de Helmholtz, a análise das suas divergências e de suas funções resposta, como função de um parâmetro intensivo como a temperatura, permite caracterizar as

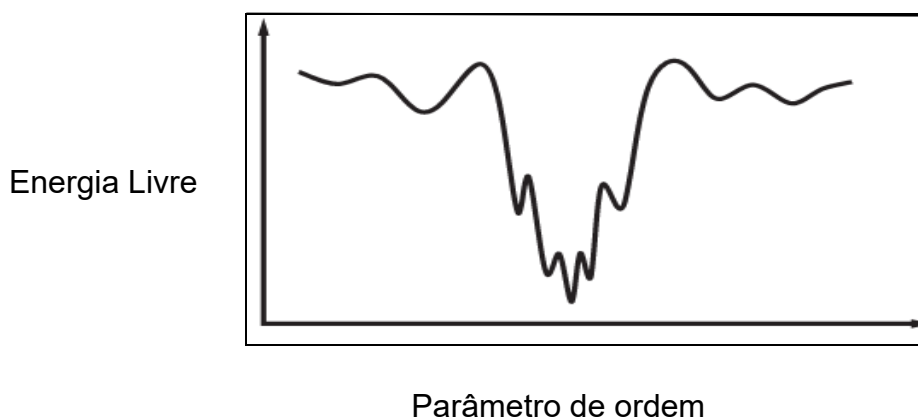


propriedades termodinâmicas das transições de fase (STANLEY, 1987).

Em nosso estudo, não nos aprofundaremos nos aspectos teóricos mais específicos das transições de fases, que constituem por si só uma vasta área de pesquisas. No entanto, é indispensável entender os fundamentos e as propriedades das separações de fases, da existência de pontos críticos e da entropia microcanônica. Isto é possível correlacionando as transições de fases termodinâmicas com suas ordens de transição. Para tal, lembramos que a entropia microcanônica  $S(E)$  em uma fase única, por exemplo de um sistema homogêneo composto por água líquida, graficamente apresenta um comportamento côncavo como função da energia interna (GROSS, 2001).

Durante o processo de *folding* das proteínas, os parâmetros relevantes a serem considerados são as cinéticas de reações químicas, bem como, as ordens de reações a uma transição de fase termodinâmica, ou como é frequentemente usado nas descrições de sistemas metaestáveis. A **Figura 9** representa um perfil de energia livre durante o processo de dobragem e pode ser descrito como um análogo de uma transição de fases, onde é possível analisar a formação de diferentes estruturas (BACHMANN, 2014).

**Figura 9: Metaestabilidade durante o *folding*.**



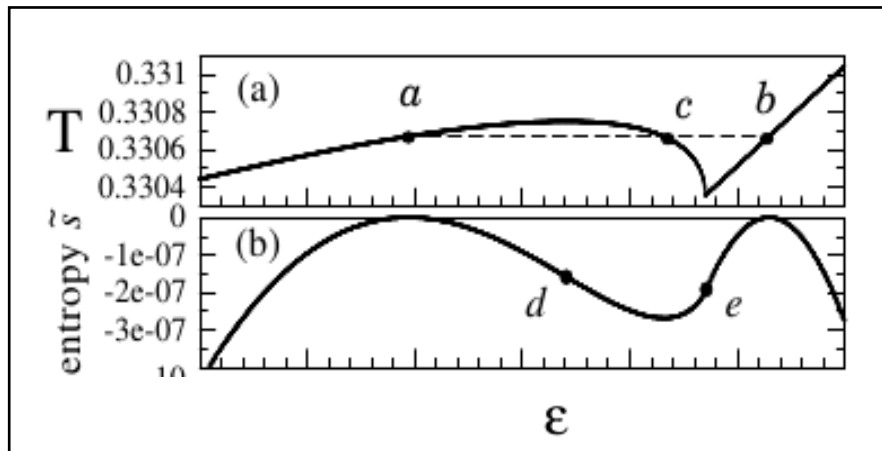
Fonte: Adaptado de Bachmann, 2014.

Em uma transição de primeira ordem, em que o aparecimento de um pico no diagrama, indica presença de fases distintas, ou seja, uma mudança brusca nas propriedades intensivas do meio. Como em um sistema composto por água líquida e gelo a 273 K, onde a entropia microcanônica apresenta um intruso convexo como

função da energia interna. Ou seja, em sistemas finitos sua curva calórica (i.e. temperatura vs energia interna) apresenta graficamente os típicos “loops de Van der Waals”, que caracterizam a presença de duas fases termodinamicamente diferentes.

Nesse caso, são necessárias as chamadas construções de Maxwell de áreas iguais, conforme visto na **Figura 10**, para determinar teoricamente as temperaturas críticas de transições, a partir de uma descrição microscópica (GROSS, 2001). Note na **Figura 10**, o diagrama que representa a curva calórica da temperatura em função da energia interna ( $T \times \epsilon$ ). Verifique na construção de Maxwell à temperatura crítica  $T_c$  feita entre os pontos “a”, “b” e “c” (i.e., “S-bend”) para corrigir os chamados loops de Van der Waals. O segmento horizontal ( $a \leftrightarrow b$ ) é construído para corrigir a curva, o que possibilita extrair propriedades termodinâmicas como calor latente de transição ( $\Delta L$ ), que é justamente o comprimento deste seguimento.

**Figura 10: Curvas calórica ( $T \times \epsilon$ ) e da entropia microcanônica ( $s$  vs  $\epsilon$ ) ao longo de uma transição de fase.**



Fonte: Adaptado de Frigori, 2010.

Ainda da análise e interpretação do diagrama, observa-se que, entre  $c \rightarrow b$  há uma variação brusca na temperatura no ponto  $c$ , simultaneamente com aumento da energia. Em seguida a temperatura aumenta constantemente passando pelo ponto  $b$ , pseudo-crítico, caracterizando que há coexistência de fases distintas. Em nosso estudo utilizaremos o princípio desta metodologia para estabelecer nossos resultados.

#### 4.4 TEORIA CLÁSSICA DE NUCLEAÇÃO

Para nosso estudo, também precisaremos entender, os conceitos e fundamentos termodinâmicos de sistemas homogêneos e heterogêneos, cinéticas de reações, bem como, pontos críticos que são revelados nas transições de fases. Para tanto, recorreremos à teoria clássica de nucleação como descrita em (FOKIN, YURITSYN e ZANOTTO, 2005). Para melhor entender, consideremos como uma analogia ao comportamento teórico da agregação de proteínas a cristalização do vidro. Este material é caracterizado como um sólido amorfo não cristalino, em cujo processo de formação, passa da fase líquida para a sólida, por meio de uma transição de fase termodinâmica de primeira ordem (i.e. com presença de calor latente), chamada de transição vítrea. Sua cristalização se dá durante o processo de resfriamento, em que há redução da energia interna e ocorre aumento gradativo de sua viscosidade até tornar-se um sólido amorfo, um agregado de sílica. A nucleação pode ser induzida em ensaios *in vitro* com interfaces preexistentes, fases previamente nucleadas e defeitos superficiais em soluções heterogêneas, durante o processo de cristalização e derretimentos do vidro (FOKIN, YURITSYN e ZANOTTO, 2005).

Em nossa proposta, esse fenômeno termodinâmico de transição de fase, será analisado teoricamente e entendido como o processo de transição entre a fase solvatada da insulina com a amilina, e os seus análogos de interesse (representando uma “fase líquida”). Enquanto a agregação desses componentes produziria uma fase agregada (equivalente a um “sólido amorfo”). A matemática que descreve ambos os sistemas (vidro ou proteínas) durante suas respectivas transições de fase de primeira ordem, é indiferente ao material. Assim, analisaremos semelhantemente as propriedades das transições de fases de agregação. Em proteínas, o análogo à transição de fase vítrea é estudado por métodos empíricos cineticamente através da evolução de indicadores de fluorescência (ThT). Como em (SILVA e LIMA, 2017), que foi utilizado uma solução tamponada de Pramlintide com tioflavina e verificou-se o seu tempo de agregação e precipitação da solução, evidenciado por meio da evolução temporal da fluorescência da solução.

## 5 SIMULAÇÕES COMPUTACIONAIS

Segundo (PADGEN, 1991) “simulação é o processo de projetar um modelo computacional de um sistema real e conduzir experimentos com este modelo com o propósito de entender seu comportamento e/ou avaliar estratégias para sua operação”. Em nosso caso, seria a simulação de um experimento *in vitro* sem necessidade de ir a um laboratório, obtendo resultados precisos pelo poder computacional, além de evitar gastos dispendiosos com itens consumíveis como reagentes e vidrarias, como utilizados em métodos de cristalografia de difração de raio x e ressonância magnética nuclear por exemplo.

### 5.1 MODELOS DE SIMULAÇÕES

Atualmente existem vários modelos para análises de *foldings* de proteínas, sendo que alguns são mais promissores. Alguns deles são as ferramentas de bioinformática de *design*, sequências e estruturação de proteínas que possibilita possíveis combinações de receptores e ligantes em agregados, geralmente com base na verificação de banco de dados de acordo com determinados critérios e grandezas inseridas (BACKMANN 2014 pg 15). A seguir, alguns exemplos de métodos de simulações:

#### 5.1.2 Simulações de Monte Carlo

Devido a impossibilidade de se simular um sistema físico no limite termodinâmico, ou seja, com um número de componentes microscópicos tendendo ao infinito ( $N \rightarrow \infty$ ), é necessário estudar as médias estatísticas feitas em sistemas finitos. Na abordagem microcanônica isso é possível, pois a entropia microcanônica  $S(\epsilon, N, V)$  pode ser rigorosamente definida para um sistema com volume ( $V$ ), energia interna ( $\epsilon$ ) e

número de componentes (N) finitos, através da célebre fórmula de Boltzmann:

$$S(\epsilon, N, V) = k_B \ln \Omega(\epsilon, N, V) \quad (1)$$

Onde  $\Omega(\epsilon, N, V)$  é o número de microestados estatísticos associados a um macroestado termodinâmico, onde  $k_B$  é a constante de Boltzmann introduzida na equação para dar uma dimensão física para entropia (S).

Para implementação desse cálculo é possível utilizar métodos numéricos, como os algoritmos de Monte Carlo (NEWMAN e BARKEMA, 1999). Em particular, em nossos estudos empregaremos técnicas Multicanônicas como em (FRIGORI, 2010) e implementadas no *software* SMMP (MEINKE *et al.*, 2008).

Esta metodologia consiste em amostrar configurações de proteínas sorteadas aleatoriamente segundo um peso estatístico  $\omega_{muca} = (E_k)$  e armazenar suas respectivas energias  $E_k$ , a fim de construir iterativamente histogramas energéticos  $H_{muca}(E)$ . Esses pesos estatísticos, ditos multicanônicos, são alterados sistematicamente à medida que a simulação dos sistemas evolui, por meio de uma relação de recorrência (BERG, 1996):

$$\omega_{muca}^{n+1}(E) = e^{-S_{muca}^{n+1}(E)} = e^{\frac{\omega_{muca}^n(E)}{\hat{H}_{muca}^n(E)}} \quad (2)$$

que ao final do processo iterativo ( $n \gg 1$ ) permitem determinar a entropia microcanônica  $S_{muca}(E_k)$ , e com ela extrair a termodinâmica do sistema.

A saber, segundo (GROSS, 2001), podemos determinar observáveis termodinâmicos, como o inverso da temperatura estatística ( $\beta^{-1}$ ) através da derivada da entropia com relação à energia interna:

$$k_B \beta(E) \equiv T^{-1}(E) = \frac{\partial S}{\partial E} \quad (3)$$

Bem como, o calor específico a volume constante que é calculável através da relação:

$$c_v(E) \equiv \frac{dE}{dT} = - \left( \frac{\partial S}{\partial E} \right)^2 \left( \frac{\partial^2 S}{\partial E^2} \right)^{-1} \quad (4)$$

e a energia livre de Helmholtz (F), a uma temperatura fixa  $T_c$ , obtida em:

$$F(E) = E - \left( \frac{\partial S}{\partial E} \right)^{-1} \Bigg|_{E=E(T_c)} S(E) \quad (5)$$

## 5.2 CAMPO DE FORÇA ECEPP/3

Os campos de forças existentes foram desenvolvidos de maneira independente e com todos os conjuntos de parâmetros específicos para descrever as ligações de hidrogênio, ângulos e comprimentos de ligações. A escolha do campo de força depende, em grande parte, do sistema a ser analisado e das propriedades que serão investigadas. No caso de sistemas biomoleculares, os campos de força mais utilizados são *CHARMM*, *GROMOS*, *AMBER*, *OPLS* (NAMBA, da SILVA e SILVA, 2018) e (MEINKE, *et al.*, 2008) entre outros *all atoms* como *ECEPP/3*, que optamos por utilizar em nosso estudo.

O ECEPP/3 possibilitou-nos avaliar, os movimentos físicos dos átomos e moléculas aos quais associam-se potenciais de interações, que coordenam seus movimentos. A descrição do conjunto de todas as interações entre esses átomos é denominada de campo de força. Na implementação do ECEPP/3, as interações entre pares de átomos ligados não-covalentemente, são descritas por potenciais compostos pelos termos de Van der Waals e eletrostático, representados respectivamente pelos potenciais Coulombiano ( $E_C$ ), Lennard-Jones( $E_{LJ}$ ), de Hidrogênio ( $E_{hb}$ ) e Torção ( $E_{tor}$ ), onde suas descrições podem ser caracterizadas matematicamente como a soma dos componentes da energia intramolecular e seus componentes eletrostáticos (MEINKE, *et al.*, 2008) (EISENMENGER *et.*, *al*, 2001) como segue:

$$E_{\text{Intra}} = E_C + E_{LJ} + E_{\text{hb}} + E_{\text{TOR}} \quad (6)$$

ou seja:

$$E_{\text{Intra}} = \sum_{(i,j)} \frac{332q_iq_j}{\epsilon r_{ij}} + \sum_{(i,j)} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{(i,j)} \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \sum_l U_l [1 \pm \cos(\eta_l \psi_l)] \quad (7)$$

onde  $q_i$  é a carga parcial do átomo  $i$ ,  $r_{ij}$  é a distância entre os átomos  $i$  e  $j$ ,  $\psi_l$  é o  $l$ -ésimo ângulo de torção e a profundidade do potencial de torção e o número de seus mínimos são dados respectivamente por  $U_l$  e  $\eta_l$ . As energias são medidas em kcal/mol. Os parâmetros  $A_{ij}$ ,  $B_{ij}$ ,  $C_{ij}$  e  $D_{ij}$  que dependem dos tipos atômicos, bem como  $U_l$  e  $\eta_l$ , que são determinados a partir das características dos peptídeos (EISENMENGER *et. al.*, 2006). Para incluir os efeitos do solvente no *fold*ing, inclui-se um termo energético baseado na área de superfície acessível da proteína pelo solvente:

$$E_{\text{sol}} = \sigma_i A_i \quad (8)$$

Onde  $A_i$  é a superfície exposta do  $i$ -ésimo átomo e  $\sigma_i$  é um parâmetro de solvatação determinado empiricamente a partir das energias livres da solução de pequenas moléculas. Além disso, pode-se considerar interações interproteicas entre cadeias multipéptídicas, que levam à agregação, o que é de fundamental interesse para nossos estudos de dimerização. Para esse fim, foi adicionado um termo inter-molecular extra  $E_{\text{Inter}} = E_C + E_{LJ} + E_{\text{hb}}$  representando as interações entre átomos intermoleculares de modo que a energia total do sistema multi-molecular solvatada se torne:

$$E_{\text{total}} = E_{\text{intra}} + E_{\text{inter}} + E_{\text{sol}} \quad (9)$$

Os campos de forças existentes foram desenvolvidos de maneira independente e com todos os conjuntos de parâmetros específicos para descrever as ligações de hidrogênio, ângulos e comprimentos de ligação. O ECEPP/3 em comparação com outros campos de força, caracteriza-se por sua geometria fixa, onde, as orientações

espaciais e os ângulos dos átomos podem ser alteradas, mas a distância entre os átomos em ligações covalentes permanece fixa.

Por isso, ocorrem menos etapas nos processos de atualizações moleculares do programa, o que o torna mais rápido, simplifica as superfícies da energia potencial molecular, permitindo a produção de grandes volumes de dados estatísticos, o que é conveniente para estudos termostáticos. Porém, essa abordagem pode ser menos precisa em relação aos campos com de geometria flexível. Nossa abordagem segue o protocolo validado em (FRIGORI, 2017) que otimizou a pramlintide utilizando isoformas de amilina.

### 5.2.1 Modelos de Solvatação

São modelos computacionais que representam o comportamento de fases condensadas solvatadas. Modelos de solvatação permitem simulações e cálculos termodinâmicos aplicáveis a reações e processos que ocorrem em solução. Podemos definir a solvatação como a combinação dos íons das moléculas de um soluto com as moléculas do solvente, ou seja, o processo pelo qual as moléculas do soluto são envolvidas pelo solvente para formar um novo composto.

Modelos de solventes têm sido vastamente testados e revisados na literatura científica (SPOEL *et al.*, 2009). Os vários modelos geralmente podem ser divididos em duas classes, modelos explícitos e implícitos, cada qual tem suas próprias vantagens e desvantagens.

- **Modelos Implícitos:** são computacionalmente eficientes, rápidos, mas imprecisos por que o solvente é descrito como um “contínuo” que interage com o soluto apenas pelo termo energético de “Área Exposta” como por exemplo: SCH2.
- **Modelos explícitos:** são computacionalmente menos eficientes, mais lentos, porém, bem mais precisos, visto que representam moléculas individuais de solvente (água), que interage com soluto por meio de campos de força bastante detalhados. Explicam dados experimentais com bastante precisão a exemplo TIP4P-EW. Não funcionam em conjunto com métodos de Monte Carlo.



O campo de força atômico pode ser complementado por modelos de interações das proteínas com solventes cada vez mais precisos, permitindo previsão de estruturas proteicas com dimensões abaixo de 3,5 Å. Esse equilíbrio característico de simplicidade e precisão, favoreceu a incorporação do ECEPP/3 a um modelo de solvatação determinado empiricamente a partir de energias livres em solução com pequenas moléculas, o SCH2.

### 5.2.2 Modelagem da Solvatação Com PBEQ-Solver

O processo de solvatação requer um gasto de energia, que é empregada no rompimento das ligações entre os íons do soluto e das interações das moléculas do solvente, sendo endotérmica estas duas primeiras etapas do processo. Em contrapartida, há uma liberação de energia devido à formação das interações do soluto com o solvente, constituindo a fase exotérmica. A solvatação está constantemente presente em diversos processos que ocorrem no organismo humano. Inúmeras substâncias orgânicas e também inorgânicas são transportadas em meio aquoso através desse mecanismo de dissolução, como é o caso de proteínas, íons, gases, nutrientes, hormônios e sais.

O PBEQ-Solver realiza uma modelagem de solvatação operada por conexão remota na ([CHARMM-GUI, 2020](#)): Uma plataforma para mecânica e dinâmica macromolecular com ferramentas versáteis de análise e manipulação de coordenadas atômicas e trajetórias dinâmicas. O *PBEQ-SOLVER* calcula e visualiza o potencial eletrostático de moléculas, resolvendo a equação de Poisson-Boltzmann (PB) que rege o potencial elétrico de cargas distribuídas homoganeamente no meio aquoso, separadas por distâncias ( $r$ ) que determinará estatisticamente o potencial elétrico gerado pelos íons dispersos na solução (BUTT, GRAF e KAPPL, 2006).

A plataforma permite calcular potencial eletrostático, energia de solvatação, energia de interação eletrostática proteína-proteína e o pKa de um resíduo em ambientes aquosos de solvente e membrana. Essas variáveis são obtidas pelo envolvimento total das partículas, moléculas e íons de um soluto com a água, ou seja,

quando ocorre a dissolução, essas quantidades são numericamente calculadas *on line*, sendo necessário fornecer como dados de entrada apenas um arquivo PDB.

Em nosso estudo optamos pelo cálculo da energia de solvatação, dimensionada em kcal/mol, ou seja, a quantidade de energia liberada durante a reação de dissolução, por unidades molares da solução gerada. No PBEQ-solver os usuários podem ainda especificar diversos parâmetros físicos para cálculos de (PDBs) e visualizar interativamente o potencial eletrostático na superfície acessível por solvente, bem como, os contornos do potencial isoeletrostático ([CHARMM-GUI, 2020](#)).

### 5.3 SOFTWARES DE VISUALIZAÇÃO E MODELAGEM MOLECULAR

São utilizados para plotar, manipular e visualizar as estruturas que são codificadas em arquivos de formatos com extensão FASTA ou PDBs ([RCSB Protein Data Bank, 2020](#)) conforme já vistas. As pesquisas e os resultados obtidos são mais bem apresentados e interpretados quando da utilização dessas ferramentas. Existem alguns tipos de *softwares* mais usados para essa finalidade. Alguns, destes inclusive usamos em nossas análises:

- **RasMOL:** Programa de visualização gráfica molecular, usado para visualizar macromoléculas e preparar imagens com qualidade de publicação ([RasMOL, 2020](#)).
- **PyMOL:** Visualizador molecular de alta performance com suporte para animações e renderização de alta qualidade. Com rotinas de cristalografia e outras atividades moleculares gráficas usuais, sendo compatível com sistemas operacionais *Windows, Linux* e *macOS* ([PyMOL, 2015](#)).
- **Jmol:** É um visualizador de estruturas químicas em 3D. Pode ser utilizado como uma ferramenta de ensino, ou para pesquisas em química e bioquímica. É um *software* livre e aberto, escrito em Java e por isso, o *software* roda em *Windows, MacOS X, Linux* e *Unix*. Nele, as moléculas podem ser exibidas como modelos de "bastão e bola". Suporta uma ampla gama de formatos de

arquivos moleculares, incluindo o formato *Protein Data Bank* (PDBs), *Crystallographic Information File* (CIF), *MDL Molfile* (mol), e *Chemical Markup Language* (CML). Ainda possui recursos para química, biomoléculas, cristais e outros materiais ([JMOL, 2015](#)).

- **VMD** (*Visual Molecular Dynamics*): um programa de computador para modelagem molecular e visualização, sendo empregado como ferramenta para visualização e análise dos resultados de simulações de dinâmica molecular, mas também pode trabalhar com dados volumétricos, sequência e objetos gráficos arbitrários. Os usuários podem executar seus próprios *scripts* no VMD, já que este inclui embutidos interpretadores *Tcl* e *Python*. É executado em *Unix*, *MacOS X* e *Windows* ([VMD, 2020](#)).

## 6 MATERIAIS E MÉTODOS

Recentemente, o aumento do poder computacional tem permitido realizar simulações com grandes volume de dados. Os *softwares* também têm sido aprimorados e constantemente novos algoritmos e pacotes específicos são criados através linguagens de programação para gerar, manipular e ajustar os dados das simulações. Em nossa metodologia empregamos o pacote computacional de código-fonte aberto chamado de SMMP de MEINKE, *et al.*, (2008).

Os dados de nosso estudo foram simulados por meio de um SSH (*secure shell*), que atua como uma ponte de conexão remota do nosso computador pessoal ao supercomputador Santos Dumont que está alocado ao LNCC (Laboratório Nacional de Computação Científica). A conexão acontece entre o servidor e o computador do usuário, de forma segura, sem possibilidade de interceptação do pacote de dados que está sendo transmitido.

Também o pacote *PHAST*, um *software* escrito em linguagem *Fortran*, com extensões MPI e CUDA, para realizar as simulações. A metodologia de FRIGORI (2017) permitiu através dos comandos do módulo *ANALYST* plotar os gráficos dos resultados de nosso estudo que foram analisados diretamente dentro de módulo de Termodinâmica Estatística Microcanônica. Computamos a entropia, calor específico e energias livres, fazendo *outputs* em programas como *Xmgrace* e *Gnuplot* para plotar gráficos e “mapas de calor”. Utilizamos o *RasMol* e *VMD*, para modelagem de estruturas proteicas. O *Qtplot* para manipulação e ajustes de dados, todos os programas sendo executados no sistema operacional *LINUX 18.04 LTS*.

Utilizamos como sistemas-alvo insulina, amilina e seus análogos, para estudar suas transições de fase estruturais de primeira ordem de agregação. Selecionamos um único dímero da estrutura do PDB 4EWZ, comum em drogas farmacêuticas, como um modelo básico para insulina humana de ação regular.

A sequência primária de ambas as cadeias de insulina A e B é mostrada na **Tabela 2**. No entanto, nenhuma estrutura 3D para os peptídeos *slAPP* ou *slAPP<sup>+</sup>* está atualmente disponível no PDB. Assim, as sequências de aminoácidos de *hlAPP*, *slAPP*

ou sIAPP<sup>+</sup> vistas na **Tabela 2** foram empregadas para auto-inicialização dos *softwares* de simulações como cadeias linearmente conformadas.

**Tabela 2: Sequência de aminoácidos FASTA para Amilina humana, Pramlintide, S20R-Pramlintide com as respectivas mutações em vermelho e as cadeias de Insulina humana de ação regular (4EWZ).**

PEPTÍDIO	SEQUÊNCIA
Amilina humana	KCNTATCATQRLANFLVHSSNCFGAILSSTNVGSNTY
Pramlintide	KCNTATCATQRLANFLVHSSNCFG <b>PILPP</b> TNVGSNTY
S20R-Pramlintide	KCNTATCATQ.RLANFLVHS <b>R</b> NNCFG <b>PILPP</b> TNVGSNTY
4EWZ (cadeia A)	GIVEQCCTSICSLYQLENYCN
4EWZ (Cadeia B)	FVNQHLCGSHLVEALYLVCGERGFFFTPKT

Fonte: Dados da pesquisa do autor.

Empregamos simulações multicanônicas para analisar o tempo de agregação e a solubilidades das co-formulações. O *setup* computacional das simulações de Monte Carlo é semelhante à FRIGORI (2017), incluindo assim o campo de forças ECEPP/3 e fazendo uso de solvatação implícita no modelo SCH2 como implementada no pacote SMMP 3.0 (2008).

A constante de Boltzmann utilizada é fixada como  $k_B = 1.987 \times 10^{-3}$  kcal / mol / K. Também utilizamos como container do sistema, uma caixa cúbica com lados de 200 Å para garantir uma mistura equimolar dos peptídeos, em uma concentração de aproximadamente 200 μM. Esta é uma concentração bastante próxima do nominalmente preparado em experimentos *in vitro* ensaiados por WANG *et al.*, (2014). Para amostrar esses sistemas utilizamos um algoritmo multicanônico na versão de

BERG e NEUHAUS, (1991). Tal metodologia MUCA, anteriormente já mencionada, relaciona os pesos estatísticos  $\omega_{1/\Omega}$  à entropia  $S(E)$  e às densidades de estado dos sistemas  $\Omega(E)$  pela relação em que a temperatura inversa  $\beta(E)$  e a energia livre  $\bar{\alpha}(E)$  são expressas na Eq. (10).

$$\omega_{1/\Omega} = \frac{1}{\Omega} = e^{-S(E)} = e^{-\beta(E)E + \bar{\alpha}(E)} \quad (10)$$

O cálculo da Eq.(10) é feito com uma técnica iterativa, em que se constroem histogramas  $H_{muca}(E)$ , compostos por energias internas  $E_k \in [E_0, \dots, E_{max}]$  durante a evolução de Monte Carlo. O processo iterativo para quando a eq. (2) converge e, portanto, fornece uma equação numericamente estável para a entropia escrita em termos dos parâmetros MUCA, i.e.  $\{\beta_k, \alpha_k\}$ .

$$S_{muca}(E_k) = \beta_k E_k - \alpha_k \quad (11)$$

e em energia livre de Helmholtz, à temperatura crítica  $T_c$  via Eq. (5), como em (FRIGORI, 2017). A agregação de proteínas exemplifica um fenômeno de transição de fases de primeira ordem GROSS (2001). Assim, as temperaturas de transições pseudo críticas podem ser calculadas através das construções de Maxwell das curvas calóricas  $\beta(E) \times E$  nas vizinhanças dessa região (destacada pelo aparecimento de “S-bends”). A respectiva região energética associada à construção de Maxwell corresponde ao calor latente ( $\Delta L$ ) dado por:  $\Delta L = E_{agg} - E_{frag}$ , em que temos as energias de agregação  $E_{agg}$  e a de fragmentação  $E_{frag}$  como situações limite GROSS (2001).

Limitamos nossa análise às temperaturas entre 224,8K até 647K, ou seja, entre os pontos de nucleação homogêneo do gelo e o ponto crítico da água pois além deste intervalo não pode haver água líquida.

Considerando que, como os ensaios de agregação normalmente empregam tioflavina SILVA *et al.*, (2016) para monitorar a cinética de formação de amilóides, o tempo requerido para esta reação entrar em estado estacionário é conhecido como

**time-lag** ( $\tau_{lag}$ ). Baseado na teoria clássica da nucleação (CNT) FOKIN, YURITSYN e ZANOTTO (2005), podemos estimar estes tempos partindo de valores termodinâmicos simulados computacionalmente.

Segundo a hipótese de nucleação homogênea, a força-motriz termodinâmica que produz transições de nucleação é proporcionada pela variação entre as energias livres das fases fragmentada e agregada  $\Delta G_V$ . Sendo assim a uma temperatura  $T$ , abaixo da temperatura de agregação  $T_c$ , a aproximação de Hoffman implica:

$$\Delta G_V(T) = \Delta H_V \left(1 - \frac{T}{T_c}\right) \frac{T}{T_c}, \quad (12)$$

onde  $\Delta H_V$  é chamado de entalpia de derretimento do cristal, que é igual ao calor latente de agregação  $\Delta L$ . Ainda, de acordo com a CNT o  $\tau_{lag}$  pode ser calculado pela relação

$$\tau_{lag} = \frac{\eta}{\Delta G_V^2} \exp\left(\frac{\Delta G_D}{k_B T}\right) \quad (13)$$

onde  $\eta$  é uma constante que depende de especificidades do sistema, enquanto que  $\Delta G_D$  é a barreira de energia livre de (Helmholtz) entre a fase agregada e não-agregada:

$$\Delta G_D \simeq \Delta F(E)|_{E=E^*(T_c)} \quad (14)$$

e esta barreira  $\Delta F$  é calculada utilizando a Eq (5). Para compararmos as estimativas teóricas com valores aferíveis experimentalmente, calculamos a razão  $RT \equiv \tau_{2s} / \tau_{2h}$  dentre os respectivos *time-lags* de agregações da sIAPP ( $\tau_{2s}$ ) e ( $\tau_{2h}$ ) da hIAPP como em FRIGORI (2017). Esperamos ainda que quanto maiores os *time-lags* de agregação, maiores serão as estabilidades moleculares e consequentemente a solubilidade, cuja componente eletrostática é dada por  $\Delta G_{elec} = \sum_k q_k \Delta \phi_{vs}$ . Essa perspectiva é coerente com a recente análise de energia livre de solvatação  $\Delta G_{solv}$  de mutantes monoméricos de amilina realizada por ALVES e FRIGORI (2019).

Em nosso modelo contínuo de solvatação IM, BEGLOV e ROUX (1998), a energia-livre total  $\Delta G_{sol} = \Delta G_{np} + \Delta G_{elec}$  é composta por uma parcela não-polar ( $\Delta G_{np}$ ) e uma parcela eletrostática ( $\Delta G_{elec}$ ), respectivamente. Usualmente a parcela não-polar é expressa como o produto da superfície do soluto por um coeficiente  $\gamma$ . O componente eletrostático, por sua vez, é escrito em termos da carga  $q_k$  do K-ésimo átomo inserido no vácuo  $\epsilon_v$  e no solvente  $\epsilon_s$ . A energia potencial eletrostática foi obtida da equação de Poisson-Boltzmann utilizando o servidor online PBEQ-solver, que também nos permitiu calcular a energia livre de solvatação, conforme JO, *et al.*; (2008). Particularmente, configuramos as constantes dielétricas do solvente  $\epsilon_s = 78.5$  e da proteína  $\epsilon_p = 2.0$  bem como concentração de sal de 150 mM.

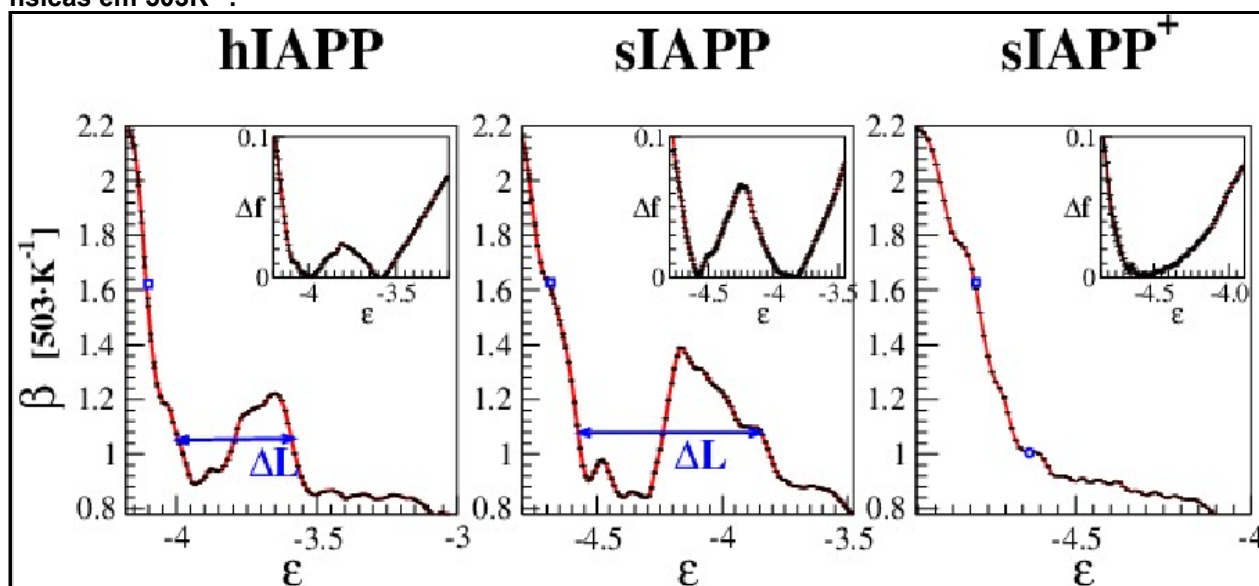


## 7 RESULTADOS E DISCUSSÕES

A partir dos parâmetros MUCA  $\{\beta_K, \alpha_K\}$ , obtidos iterativamente a partir das simulações de agregação feitas com o SMMP, utilizamos o pacote PHAST desenvolvido em FRIGORI (2017) para as determinações termodinâmicas das quantidades Eq.(12), Eq. (13) e Eq. (14). Assim, fomos capazes de extrair a termoestatística microcanônica do processo de heterodimerização de insulina com amilina (hIAPP) e também com seus análogos biotecnológicos (i.e. sIAPP ou sIAPP<sup>+</sup>).

Enfatizamos que as curvas calóricas [ i.e.  $\beta(\epsilon) \times \epsilon$  ] e a energias livres de Helmholtz ( $\Delta f$ ) são exibidas na **Figura 11**, e aparecem como função da energia interna por resíduo ( $\epsilon$ ), pois no ensemble microcanônico a energia interna é o parâmetro de controle e não a temperatura. Nesta mesma figura, vemos as regiões canônicas termodinamicamente instáveis, os *S-bends*, onde fizemos as construções de Maxwell conforme visto com seta dupla, horizontal e azul.

**Figura 11:** Curvas calóricas para análogos de amilina em associação com insulina humana de ação regular. Insetos: mostram as respectivas Energias livres de Helmholtz ( $\Delta f$ ) avaliada na temperatura pseudo-crítica ( $T_{agg.}$ ) como função da energia interna ( $\epsilon$ ). As construções de Maxwell, segmentos horizontais azuis, denotam os calores latentes canônicos de heterodimerização ( $\Delta L$ ). As unidades de energia em  $\epsilon$ ,  $\Delta f$  e  $\Delta L$  estão normalizados para kcal/mol/resíduo e as temperaturas físicas em  $503K^{-1}$ .



Fonte: Dados do autor

Para obter os calores latentes de agregação ( $\Delta L$ ) à temperatura pseudo-crítica ( $T_C^{-1} = k_B \beta_C$ ). As alturas das respectivas barreiras de energias livres de Helmholtz ( $\Delta F$ ) são mensuráveis através dos insertos dos gráficos na **Figura 11** e mostram, explicitamente, que enquanto hIAPP e sIAPP agregam-se em uma transição de primeira ordem a sIAPP<sup>+</sup> é imune à agregação ou seja, não apresenta uma barreira em  $\Delta F$ . Conforme a **Tabela 3** mostramos os valores numéricos das análises termodinâmicas. as energias de configuração mais estáveis são denominadas “*Ground states*” ( $E_{ground}$ ), as energias livres de solvatação ( $\Delta G_{solv.}$ ), Helmholtz ( $\Delta F$ ) e o calor latente (L) são medidos em kcal / mol. Evidentemente, as co-formulações de insulina com hIAPP e sIAPP apresentam transições de fase estruturais de primeira ordem (i.e. agregam-se) em altas temperaturas. Ou seja, a temperaturas fisiológicas ambos sistemas são altamente propensos à heterodimerização e portanto, formam agregados. Por sua vez, sIAPP<sup>+</sup> não produz agregados em coformulação com insulina por isso, não tem uma  $T_{agg}$ .

**Tabela 3: Resultados termodinâmicos para a heterodimerização dos análogos de amilina com a insulina humana de ação regular. Detalhes no texto.**

PROTEÍNA	$T_{agg}$ (K)	$E_{ground}$ (Kcal/mol)	$\Delta G_{solv.}$ (Kcal/mol)	L (Kcal/mol)	$\Delta F$ (Kcal/mol)
hIAPP	479	-382,6	-746(24)	35,2	2,1
sIAPP	466	-423,3	-801(37)	65,1	6,0
sIAPP <sup>+</sup>	–	-436,6	-897(52)	–	–

Fonte: Dados do autor

Por outro lado, como o sistema composto por sIAPP<sup>+</sup> e insulina não apresenta transição de fase, demonstra-se que, uma elevada estabilidade físico-química é alcançada nessa co-formulação com o uso de um análogo resistente a agregação. Estes resultados confirmam evidências de experimentos em que a insulina agrega quando co-formulada com sIAPP de acordo em da SILVA e LIMA (2018), mas este efeito é melhorado quando empregam-se análogos ricos em cargas positivas como em WANG *et al.*, (2014).

Em seguida, para obtermos uma comparação cinética entre o comportamento das co-formulações propensas à agregação calculamos, através da Eq.(13) e dados da **Tabela 3**, a razão entre os lag-times  $R = \tau_s/\tau_h$  para as co-formulações envolvendo insulina com hIAPP ( $\tau_h$ ) ou com sIAPP ( $\tau_s$ ). Ao fixarmos a temperatura de referência próximo ao valor fisiológico ( $T = 300K$ ), ou seja, compatível com o praticado em ensaios experimentais SILVA e LIMA (2018), obtivemos a razão **R = 200**.

Este resultado demonstra que, muito embora sIAPP ainda seja propensa à agregação quando em co-formulação com Insulina, ela é notadamente menos amiloidogênica que a Amilina humana. Nosso resultado é compatível com experimentos que demonstraram que, sob condições apropriadas de pH em soluções tampões, é possível co-formular Pramlintide com Insulina de forma semi-estável SILVA e LIMA, (2018). Vale destacar que, como sIAPP<sup>+</sup> não agrega, seria impossível fazer um cálculo análogo pois seu *lag-time* seria infinito.

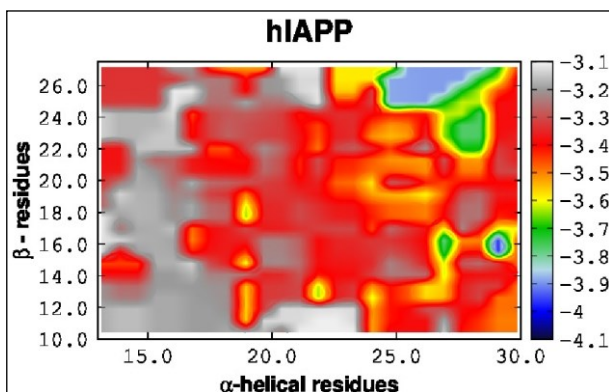
Avaliamos ainda a energia livre de solvatação para configurações simuladas de todos os sistemas a uma temperatura de 300K. Para tal, empregamos o servidor web PBEQ-solver, que resolve numericamente a equação de Poisson-Boltzmann. Obtivemos os seguintes valores da **Tabela 3** de  $\Delta G_{solv}$  -746(24) Kcal/mol (hIAPP), -801(37) Kcal/mol (sIAPP) e -897(52) Kcal/mol (sIAPP<sup>+</sup>). Mais uma vez, nossos resultados evidenciam coerentemente que sIAPP<sup>+</sup> interage mais intensamente com o solvente, logo, é mais solúvel do que Pramlintide e hIAPP, e portanto mais apta à uma efetiva co-formulação com insulina.

Para obtermos uma melhor intuição sobre os mecanismos que diferenciam as propensões à agregação dos análogos, fizemos também análises estruturais de suas co-formulações. Assim, é interessante correlacionar nossos resultados termo-estatísticos com esta análise termo-estrutural. Analisamos a prevalência de resíduos compondo hélices-alfa e folhas-beta, bem como o grau de compactação dessas estruturas, o que é fornecido pelo raio de giração  $R_{gyr}$  e que é explanado no **APÊNDICE A** e também pela distância entre os extremos moleculares  $R_{e-e}$ . De acordo com o pacote SMMP MEINKE *at al.*, (2008) estruturas com ângulos diedrais ( $\phi, \psi$ ) são classificadas como hélices no intervalo  $[-70^\circ \pm 30^\circ, -37^\circ \pm 30^\circ]$  ou como folhas-beta se  $[-150^\circ \pm 30^\circ, -150^\circ \pm 30^\circ]$ . As **Figura 12, 14, 16** apresentam os “mapas de calor” para a

ocorrência de resíduos alfa e beta, ou seja, a prevalência estrutural ao longo das configurações amostradas nas simulações como função da energia interna intensiva  $\epsilon$ , que é o parâmetro de controle do ensemble microcanônico (a temperatura canônica  $T(\epsilon)$ , relaciona-se a  $\epsilon$  por meio da curva calórica da **Figura 11**).

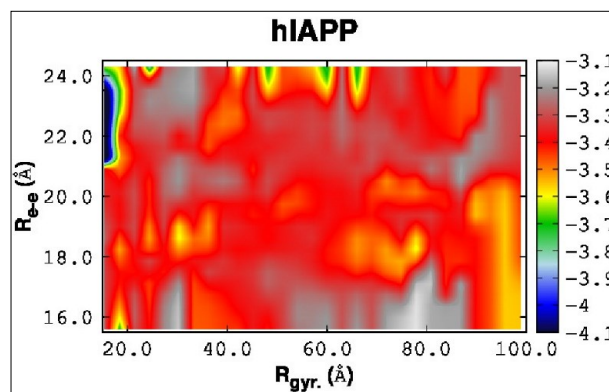
Analogamente, os gráficos das **Figuras 13, 15 e 17** representam os raios de giração  $R_{gyr}$  linear e as distâncias entre extremos moleculares  $R_{e-e}$  para os mesmos sistemas em função de  $\epsilon$ . Como consequência, ao redor da temperatura crítica de agregação, observamos nas **Figuras 12 e 14** que a co-formulação de Insulina com hIAPP ou sIAPP desenvolve uma grande prevalência de resíduos  $\beta$ , tidos como amiloidogênicos. Simultaneamente, enquanto o sistema resfria-se abaixo da temperatura fisiológica, e portanto agrega-se, essas co-formulações produzem heterodímeros compactos como mostrado na **Figura 13, 15** ou seja  $R_{gyr} < 20\text{\AA}$ . Há, portanto, clara coerência entre nossas análises termodinâmicas, **Figura 11** e nossos resultados estruturais.

**Figura 12:** Mapa de calor hIAPP com insulina



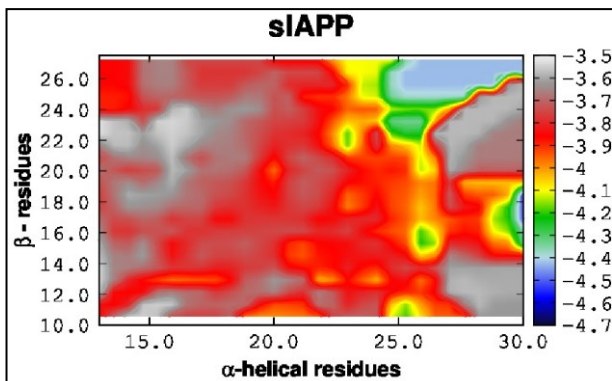
Fonte: Dados do autor

**Figura 13:** Raio de giração ( $R_{gyr}$ ) e distância entre os terminais ( $R_{e-e}$ ) da hIAPP com insulina.



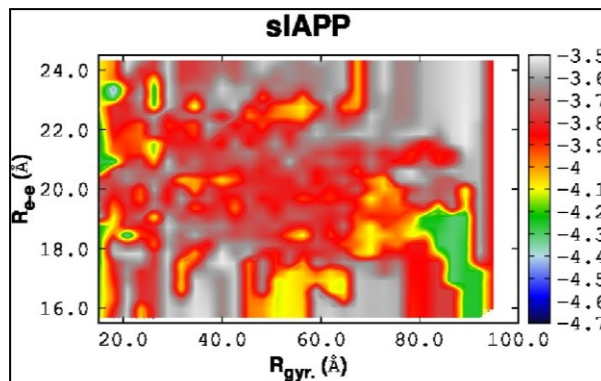
Fonte: Dados do autor

**Figura 14:** Mapa de calor siAPP com insulina



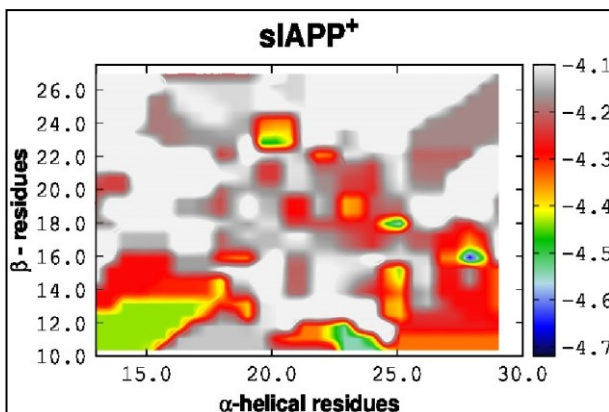
Fonte: Dados do autor.

**Figura 15:** Raio de giração ( $R_{gyr}$ ) e distância entre os terminais ( $R_{ee}$ ) da siAPP com insulina



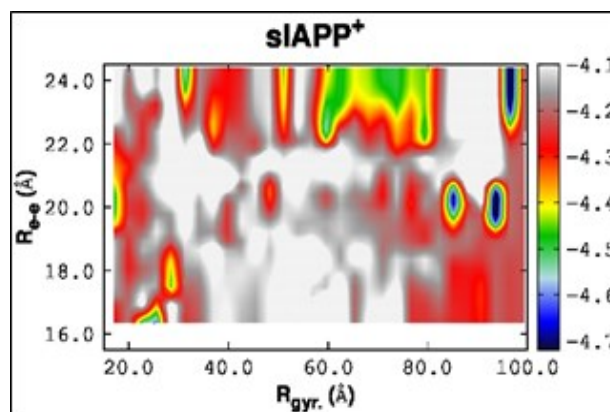
Fonte: Dados do autor.

**Figura 16:** Mapa de calor siAPP<sup>+</sup> com insulina.



Fonte: Dados do autor

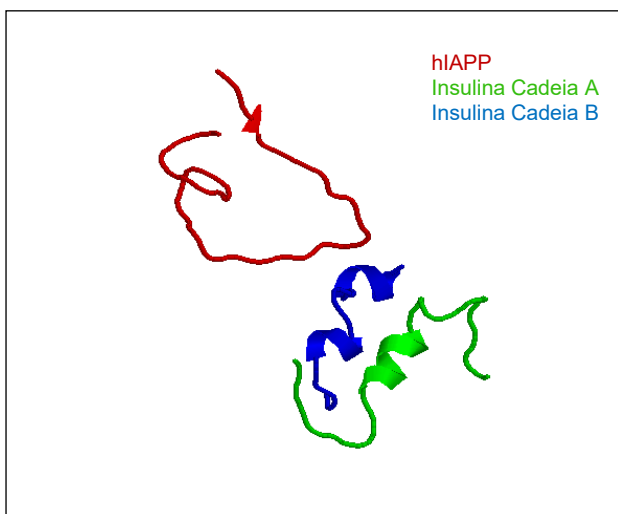
**Figura 17:** Raio de giração ( $R_{gyr}$ ) e distância entre os terminais ( $R_{ee}$ ) da siAPP<sup>+</sup> com insulina.



Fonte: Dados do autor

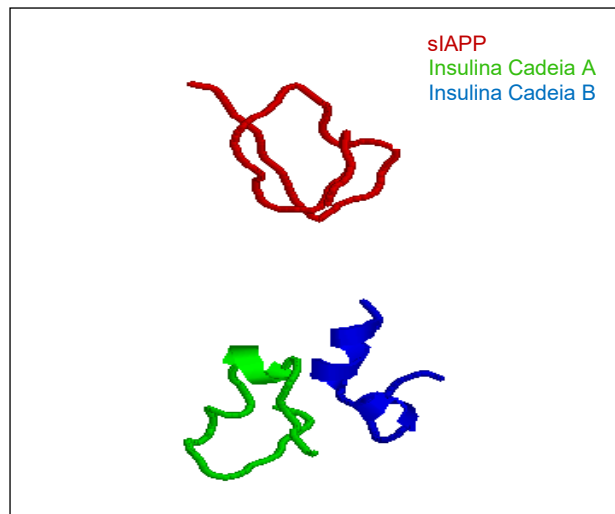
Ou seja, Amilina e Pramlintide em co-formulações com insulina agregam-se e formam núcleos agregados compactos altamente ricos em folhas  $\beta$ . Por outro lado, a co-formulação de Insulina com siAPP<sup>+</sup> não se agrega, nem exibe propensão de formar folhas  $\beta$  em temperaturas fisiológicas. Para confirmar a coerência de nossos resultados confrontamos a relação das curvas calóricas **Figura 11**, com os termogramas **Figuras 12, 13, 14, 15, 16 e 17** e com as estruturas após a agregação nas **Figuras 18, 19 e 20**.

**Figura 18: Agregação da hIAPP com insulina.**



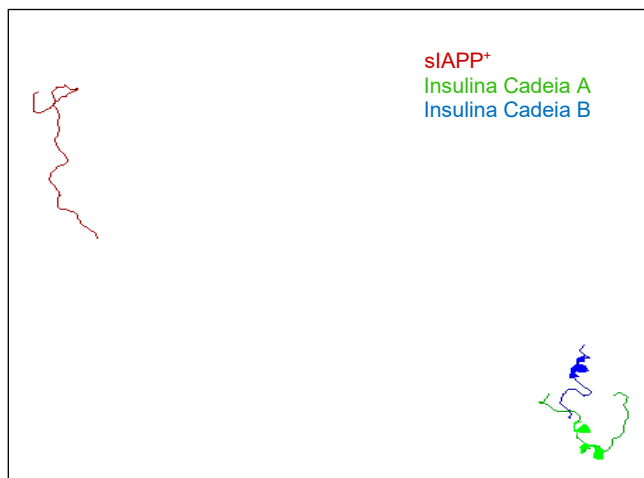
Fonte: Dados do autor.

**Figura 19: Agregação da sIAPP com insulina.**



Fonte: Dados do autor.

**Figura 20: Agregação da sIAPP<sup>+</sup> com insulina.**



Fonte: Dados do autor.

As **Figuras 18, 19 e 20** foram plotadas e estruturadas no *RasMol, 2020* utilizando configurações computacionais amostradas a 310 K. Ao compará-las confirmamos a coerência das análises anteriores. Ou seja, tanto na **Figura 18** (para insulina + hIAPP) como na **Figura 19** (insulina + sIAPP) observamos a presença de diferentes graus de agregação, enquanto que para **Figura 20** (insulina + sIAPP<sup>+</sup>) não há sinais de agregação.

## 8 CONCLUSÕES

Desenvolvemos simulações Multicanônicas de amilina humana (hIAPP) e de seus análogos biotecnológicos sIAPP e sIAPP<sup>+</sup> em co-formulação com insulina de ação regular, comumente encontrada em fármacos (código PDB: 4EWZ). Em nossos estudos computacionais empregamos o pacote SMMP de MEINKE *at al.*, (2008). Essa abordagem nos permitiu gerar dados termostatísticos que foram analisados microcanonicamente com o pacote computacional PHAST de FRIGORI (2017). Deste modo caracterizamos as transições de fases de agregação desses sistemas e as quantidades termodinâmicas acessíveis experimentalmente foram preditas, a saber: calor latente a volume constante, energia livre de Helmholtz e temperatura pseudo-crítica de agregação. Além do mais, empregamos a teoria clássica de nucleação FOKIN *at al.*, (2005) para extrair a cinética de agregação via *time-lags*.

Estes resultados foram complementados e correlacionados pelo estudo das energias livres de solvatação das co-formulações supracitadas, utilizando o PBEQ-solver, os cálculos confirmam que sIAPP<sup>+</sup> constitui um aprimoramento não-trivial da Pramlintide. Adicionalmente, nossas simulações confirmaram a presença de uma forte transição de fase de heterodimerização nas co-formulações de insulina com hIAPP ou com sIAPP. Isso demonstra que hIAPP e sIAPP agregam-se, quando co-formuladas com insulina, sendo portanto farmacologicamente inviáveis. Em suma, espera-se que o análogo da amilina humana sIAPP<sup>+</sup>, racionalmente desenhado com a aplicação de métodos computacionais, tenha elevada resistência à agregação e solubilidade FRIGORI (2017), ALVES e FRIGORI (2019).

Nosso presente estudo demonstrou que este análogo mantém estas propriedades também quando co-formulado com insulina de ação regular. Resultados semelhantes foram observados por experimentos que empregaram análogos da amilina humana que simultaneamente incorporaram Prolinas e Argininas extras WANG *at al.*, (2014), contudo, sem que uma explicação termodinâmica subjacente fosse apresentada. Dessa forma, este estudo, não só complementa e aprofunda teoricamente recentes resultados empíricos, mas também apresenta um novo potencial para insulinoterapias.

## 9 REFERÊNCIAS

- ALMEIDA, A.B. **Predição de Estrutura Terciária de Proteínas com Técnicas Multiobjectivo no Algoritmo de Monte Carlo**. 2016. 129 f. Dissertação (Mestrado em Ciência da Computação) – Instituto de Informática Universidade Federal de Goiás. Goiânia, 2016.
- ANFINSEN, C. B. **Principles that govern the folding of protein chains**. Science (New York,N.Y.). v. 181, p. 223-230, 1973.
- ALVES, N.A.; FRIGORI, R.B. **In silico comparative study of human and porcine amylin**. Jour. Phys. Chem. B. 2018.
- ALVES, N.A.; FRIGORI, R.B. **Synergistic long-range effects of mutations underlie aggregation propensities of amylin analogues**. J. Mol. Model. p.6, 2019.
- AVID N.L.; COX. M.M. **Princípios da Bioquímica de Lehninger**. ed. 6, p.1250, 2014.
- BACHMANN, M. **Thermodynamics and statistical mechanics of molecular systems**. Cambridge University Press (UK). Disponível em: [www.cambridge.org/9781107014473](http://www.cambridge.org/9781107014473) p.359, 2014.
- BERND, A. B. **Multicanonical Recursions**. Journal of Statistical Physics. v. 82, p. 323-342, 1996.
- BERG B. A.; NEUHAUS T: **Multicanonical algorithms for 1st order phase transitions**. Physics Letters B. Vol. 267, p. 254-260, 1991.
- BUTT, H.; GRAF, L.; KAPPL, M.; **Physics and Chemistry of Interfaces**. Weinheim, Germany: Wiley-VCH. 2 Ed, 373 pg, 2006.
- CABRAL, D.S.; T.R. LIMA, L. M. **Physico-chemical properties of co-formulation of insulin with pramlintide**. National Institute of Science and Technology for Structural Biology and Bioimaging (INBEB- INCT), Federal University of Rio de Janeiro, Rio de Janeiro. p. 26, 2018.



CHARMMM-GUI. **PBEQ Solver**. Disponível em: <http://www.charmmgui.org/?doc=input/pbeqsolver>. Acesso em: 15 jan. 2020.

CALLEN, H. B. **Thermodynamics and an Introduction to Thermostatistics**. ed. 2, p.485, 1985.

EISENMENGER, F.; *et al* **An enhanced version of SMMP—open-source software package for simulation of proteins**. *Computer Physics Communications*. v. 174, ed. 5, p. 422-429, 2006.

EISENMENGER F.; *et al*. **[SMMP] A modern package for simulation of proteins**. *Computer Physics Communications*. v.138, p. 192–212 193 2001.

FINEMAN M.S.; *et al*. **Properties of pramlintide and insulin upon mixing**. *Am. J. Health-Syst. Pharm.* v. 62, ed. 8, p. 816-822, 2005.

FOKIN, V.M.; YURITSYN N.S, ZANOTTO E.D. **Nucleation and Crystallization Kinetics in Silicate Glasses:Theory and Experiment**. Wiley-VCH Verlag GmbH & Co GER. p. 52, 2005.

FRIGORI, R.B. Be positive. **Optimizing Pramlintide From microcanonical analysis of amylin isoforms**. *Phys. Chem. Chem. Phys.* v.19, ed. 37, p.17, 2017.

FRIGORI, R.B. **Simulações Microcanônicas de Proteínas**. 2010, 94 f. Tese Departamento de Física e Matemática, Faculdade de Filosofia Ciências e Letras de Ribeirão, 2010.

FRIGORI, R.B. PHAST. **Protein-like heteropolymer analysis by statistical Thermodynamics**, *Computer Physics Communications*. v. 215, p. 165- 172, 2017.

GROSS, D.H.E. **The microcanonical thermodynamics of finite systems: The microscopic origin of condensation and phase separations, and the conditions for heat flow from lower to higher temperatures**. Hahn-Meitner Institute and Freie Universität Berlin, Fachbereich Physik, Glienickerstr. Berlin, Germany *Chem. Phys.* v.122, p.23, 2005.

HUANG, K. **Lectures On Statistical Physcs And Protein Folding**. Massachusetts Institute of Technology. p. 159, 2005.

IM, W.; BEGLOV, D.; ROUX, B. **Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation**. *Comput. Phys. Comm.* v. 111, p. 59-75, 1998.

LIPSKA, K.I.J.; HIRSCH, I.B.; IDDLE, M.C. **Human insulin for type 2 diabetes an active, less-expensive option**. v. 318, p. 23-24, 2017.

Jmol. **an open-source Java viewer for chemical structures in 3D**: Disponível em: <http://jmol.sourceforge.net/>. Acesso em: 15 Jan. 2020.

JO.; *et al.* **PBEQ-solver for online visualization of electrostatic potential of biomolecules**. *Nucl. Acids Res.* v. 36, 2008. Disponível em: <http://www.charmm-gui.org/input/pbeqsolver>: v. 36 2008 acesso em: 12 Jul. 2019.

JO.; *et al.* **A web-based graphical user interface for CHARMM**. *J. Comput. Chem.* Wiley Periodicals, Inc. *J Comput Chem.* v. 29, p. 1859–1865, 2008.

LANDAU, D. P.; BINDER, K. **A Guide Monte Carlo Simulations In Statistical Physics**. Cambridge. Disponível em: <http://www.cambridge.org/9780521842389>. ed.3, p. 489, 2009.

LEVINTHAL, C. J. Massachusetts Institute of Technology, Department of Biology Cambridge, Massachusetts. *Jour Chem. Phys.* n° 65, p.44, 1968.

LEVINTHAL, C. **in Mössbaun Spectroscopy in Biological Systems Proceedings**. University of Illinois Urbana-Champaign, Urbana. p. 22-24, 1968.

LUCA, S.; *et al.* **Peptide conformation and supramolecular organization in amylin brils constraints from solid-state NMR**. *Biochemistry.* V. 46, p. 18, 2007.

MEINKE, J.H.; *et al.* **SMMP v. 3.0 Simulating proteins and protein interactions in Python and Fortran**. *Computer Physics Communications.* v. 178, p. 459-470, 2008.

MELMED, S.W. **Textbook of Endocrinology**. Elsevier/Saunders, American Diabetes Society. ed. 2, p. 1920, 2001.

NAMBA, M.; SILVA, V.B; C.; SILVA, C.H.T.P. **Dinâmica Molecular: Teoria e aplicações em planejamento de fármacos**. Eclet. Química. São Paulo. v.33, 2018.

NCD Risk Factor Collaboration (NCD-RisC). **Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants**. v.387, p. 1513-1530, 2016.

NEDUMPULLY-GOVINDAN; *et al.* **Stabilizing Off-pathway Oligomers by Polyphenol Nanoassemblies for IAPP Aggregation Inhibition**. Sci. v.6, p. 1-12, 2016.

NEWMAN, M.E.J.; BARKEMA, G.T. **Monte Carlo Methods in Statistical Physics**. Disponível em: [http://itf.fys.kuleuven.be/~fppspXIII/material/Barkema\\_FPSPXIII.pdf](http://itf.fys.kuleuven.be/~fppspXIII/material/Barkema_FPSPXIII.pdf) Oxford University Press, Oxford. p. 136, 1999.

OPEN STAX. **Biology**. 2ed, p.1578, 2018. Disponível em: <https://openstax.org/details/books/biology-2e>. acesso em: 17 out. 2019.

PDB RCSB. **Protein Data Bank**. Disponível em: <https://www.rcsb.org/structure/1BW5> Acesso em: 25 Jul.

PEGDEN, C. D.; SHANNON, R. E.; SADOWSKI, R.P. **Introduction to Simulation Using SIMAN**. McGraw-Hill, NY. ed. 2, 1990.

PyMOL. **is a user-sponsored molecular visualization system on an open-source foundation, maintained and distributed by Schrödinger**. Disponível em: <https://pymol.org/2/>. Acesso em: 15 jan. 2020.

RasMOL. **Molecular Graphics Visualisation Tool**. Disponível em: <http://www.openrasmol.org/>. Acesso em: 12 jan. 2020.

SALINAS, SILVIO R.A. **Introdução à Física Estatística**. Editora da Universidade de São Paulo. ed. 2, p.466, 2005;

SCIENCE ABC. disponível em: <https://www.scienceabc.com/pure-sciences/what-are-prions.html>. Acesso em 21 Ago. 2019.

SMITH, J.M; VAN NESS C.; ABBOTT, M.M. **Introdução a Termodinâmica da Engenharia Química**. LTC. ed. 7, 2007.

STANLEY H.E. **Introduction to phase transitions and critical phenomena**. Oxford University. p. 308, 1978;

SPOEL, D. V. D.; *et al.* **Gromacs User Manual version 4.5.6**. Disponível em: [www.gromacs.org](http://www.gromacs.org). 2010.

VMD. **Visual Molecular Dynamics**. Disponível em: <https://www.ks.uiuc.edu/Research/vmd/>. Acesso em: 15. jan. 2020.

WANG H, ABEDINI A, RUZSICKA B, RALEIGH DP (2014): **Rationally designed, nontoxic, nonamyloidogenic analogues of human islet amyloid polypeptide with improved solubility**. *Biochemistry*. *Biochemistry*. v. 53, p. 5876-5884, 2014.

WESTERMARK, P.; ANDERSON, A.; WESTERMARK, G.T. **Islet Amyloid Polypeptide, Islet Amyloid and Diabetes Mellitus**. Departments of Medical Cell Biology and Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden, *Physiol* v. 91, ed.3, p. 795– 826, 2011.

## APÊNDICE A – RAIOS DE GIRAÇÃO

É a distância de um ponto na qual poderia concentrar a massa ( $M$ ) do corpo que reproduziria o momento de inércia ( $I$ ) como segue:

$$RG = \sqrt{\frac{I}{M}}$$

bem como o momento de inércia da  $i$ -ésima partícula ou molécula:

$$I_i = m_i \|\vec{r}_i\|^2$$

o que é melhor especificado na equação abaixo substituindo os termos;

$$RG = \left( \frac{\sum_i \|\vec{r}_i\|^2 m_i}{\sum_i m_i} \right)^{1/2}$$

Considerando como átomos das proteínas, o raio de giração torna-se uma forma de avaliar o grau de compactação das estruturas, além de auxiliar na verificação do enovelamento nas etapas do processo de *folding*, pois da terceira definição temos que a massa  $m_i$  do  $i$ -ésimo átomo e o  $\vec{r}_i$  é a sua posição em relação ao centro de massa da proteína (SPOEL *et al.*, 2009).