

Análise in silico da proteína beta-amilóide (A β 42) e suas mutantes Iowa (D23N) e Dutch (E22Q)

In silico analysis of beta-amyloid protein and its mutants Iowa (D23N) and Dutch (E22Q)

Gabriel Roger Santos de Oliveira¹, Rafael Bertolini Frigori²

RESUMO

Proteínas são sequências de aminoácidos que possuem atividades biológicas fundamentais. Para sua devida atividade, é necessário que esta sequência de aminoácidos conforme-se espacialmente, segundo o fenômeno de “folding molecular”, de modo a atingir um formato chamado de estrutura terciária nativa. Em modelos computacionais, utilizam-se descrições simplificadas, que retêm na modelagem apenas as interações físico-químicas supostamente mais relevantes à atividade molecular. Neste estudo, empregaremos o modelo AB, hidrofóbico-hidrofílico, para descrever o folding da proteína Beta Amilóide (A β 42) e suas mutantes (Iowa: D23N e Dutch: E22Q). Esta proteína é associada ao surgimento do Alzheimer. Tal doença surge com a formação de agregados proteicos de A β 42, neurologicamente citotóxicos, atualmente nenhuma cura é conhecida. Uma importante abordagem no tema são os experimentos in silico em que computação de alto desempenho é empregada para desvendar a dinâmica microscópica que leva ao surgimento desse tipo de patologia. Empregaremos o software PHAST para efetuar simulações computacionais do enovelamento de A β 42, e suas mutantes supracitadas, via método de Monte Carlo no ensemble Multicanônico. Com os resultados obtidos, podemos classificar o potencial agregacional de cada mutação ocorrida nas A β 42 com base nas barreiras de energias livres de enovelamento.

PALAVRAS-CHAVE: Biopolímeros, Folding molecular, Mecânica estatística, Método de Monte Carlo, Termodinâmica

ABSTRACT

Proteins are sequences of amino acids that have fundamental biological activities. For its proper activity, it is necessary for this sequence of amino acids to conform spatially, according to the characteristics of “molecular folding”, in order to achieve a format called tertiary structure that minimizes the free energy of the system. In computational models, we resort to simplified interference. In this study, we will employ the AB model, hydrophobic-hydrophilic, to describe the folding of the Amyloid Beta protein (A β 42) and its mutants (Iowa: D23N and Dutch: E22Q). This protein is associated with the onset of Alzheimer's. This disease arises with the formation of A β 42 protein aggregates, which are neurologically cytotoxic. An important approach to the topic are in silico experiments in which high-performance computing is used to uncover the microscopic dynamics that lead to the emergence of this type of pathology. We will use the PHAST software to carry out computational simulations of the folding of A β 42, and its aforementioned mutants, via the Monte Carlo method in the Multicanonical set. With the results obtained, we can classify the aggregational potential of each mutation occurring in A β 42 based on the folding free energy barriers.

KEYWORDS: Biopolymers, Molecular Folding, Monte Carlo Method, Statistical Mechanics, Thermodynamics.

INTRODUÇÃO

As proteínas são heteropolímeros, compostos por um grande número de aminoácidos, que possuem funções biológicas importantes, como por exemplo: estrutural,

¹ Bolsista do(a) CNPq. Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil. E-mail: gabrielroger@alunos.utfpr.edu.br. ID Lattes: 9313900176281101.

² Docente no curso de física 3. Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil. E-mail: frigori@utfpr.edu.br. ID Lattes: 5836878566801544.

metabólica e hormonal. Porém essas funções só são possíveis de serem executadas quando a proteína atinge sua estrutura terciária, também conhecida como “estrutura nativa”, enovelamento proteico é o nome que damos ao processo que faz com que a proteína saia de sua estrutura primária e atinja sua estrutura funcional.

Durante a conformação de proteínas podem haver erros no processo de enovelamento os quais originam as proteinopatias, doenças geradas por proteínas que foram mal enoveladas. Uma das proteinopatias mais conhecidas é o mal de Alzheimer. Essa doença ocorre pela agregação, no tecido cerebral, da proteína beta-amilóide (A β 42), neurologicamente citotóxica. Através da mecânica estatística podemos prever a capacidade agregacional da proteína (A β 42) e suas mutantes Dutch (E22Q) e Iowa (D23N).

A mecânica estatística é o ramo da física que estuda o comportamento de sistemas mecânicos macroscópicos compostos por um elevado número de entidades microscópicas a partir do comportamento destas entidades quando seus estados são incertos ou indefinidos. Como a abordagem microcanônica baseia-se na exata contagem de microestados ela é ideal no estudo de sistemas “pequenos”, como as proteínas. Justamente por isso conceitos como entropia, energia livre, calor específico e curva calórica serão de extrema utilidade para analisarmos a capacidade agregacional da proteína e suas mutantes.

MATERIAIS E MÉTODOS

Beta-amilóide: Selvagem, Dutch e Iowa

A estrutura primária de uma proteína corresponde à sequência linear dos aminoácidos unidos por ligações peptídicas.

Quadro 1 – Sequência primária das mutações

Mutação	Sequência Primária
Wild-Type	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA
Dutch	DAEFRHDSGYEVHHQKLVFFA <u>Q</u> DVGSNKGAIIGLMVGGVVIA
Iowa	DAEFRHDSGYEVHHQKLVFFA <u>E</u> NVGSNKGAIIG <u>N</u> L VGGVVIA

Fonte: PDB Protein Data Bank (2003).

Modelo AB *coarse-grained*

O estudo de sistemas complexos, como as proteínas, nos mostra vários obstáculos, dentre eles o computacional. Se considerarmos a modelagem física dos heteropolímeros que leva em consideração o maior nível de detalhamento estrutural teremos como consequência um alto custo computacional, justamente por isso nesse presente estudo utilizamos o modelo AB. Este modelo é formulado em um espaço contínuo e conta com interações de longo alcance entre os aminoácidos simplificadas, empregando-se apenas descrições efetivas entre resíduos centrados nos carbonos alfa. As descrições das cadeias laterais também são simplificadas, essas são mapeadas segundo seu caráter hidrofóbico ou hidrofílico respectivamente em monômeros do tipo A ou do tipo B.

Quadro 2 – Sequência AB das mutações

Mutação	Sequência Primária
Wild-Type	AABABBBBBBBBABBBAABAAABBABBBBBBAAABAAABBAAAA
Dutch	AABABBBBBBBBABBBAABAAAB <u>A</u> BBBBBAAABAAABBAAAA
Iowa	AABABBBBBBBBABBBAABAA <u>B</u> BBBBBAAABAAABBAAAA

Fonte: Elaborado pelos autores (2023).

Software PHAST

O software PHAST (figura 1) desenvolvido em Fortran, com extensões MPI e CUDA, utiliza o método de Monte Carlo para realizar simulações paralelas do processo de envelhecimento de cadeias heteropoliméricas únicas ou múltiplas através de modelos *coarse-grained*. Os dados resultantes podem ser analisados através de uma abordagem termodinâmica estatística microcanônica permitindo calcular a entropia, curva calórica, calor específico e energia livre do sistema.

Figura 1 – Tela inicial do software PHAST

```

151 =====
152 *****
153 *****
154 *****
155 *****
156 *****
157 *****
158 *****
159 *****
160 *****
161 =====
162 = Protein-like Heteropolymer Analysis by =
163 = Statistical Thermodynamics - v1.0 =
164 =====
165
166 HAMILTONIAN PARAMETERS:
167
168 r0.....: 1.000000
169 R.....: 1.200000
170 K_FENE.....: 0.000000
171 K_CURVATURE.: 0.250000
172 K_LJ_s(12,o): 4.000000
173 K_LJ_m(12,o): 4.000000
174 =====

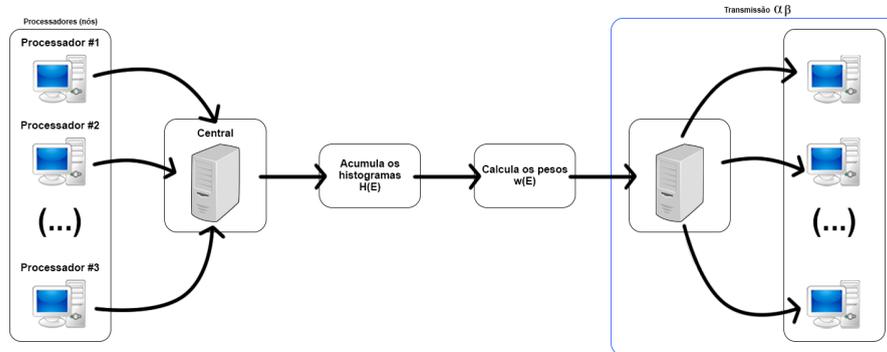
```

Fonte: Elaborado pelos autores (2023).

Computação paralela

O software PHAST emprega as extensões MPI e CUDA, associadas ao uso de recursos de paralelismo computacional, respectivamente de grão grosso e fino, para a implementação do modelo AB. O Message Passing Interface (MPI) é um padrão de comunicação amplamente utilizado na programação paralela que permite que a aplicação seja dividida em subtarefas que possam ser executadas simultaneamente em vários processadores, ou nodos, de um cluster, assim reduzindo o tempo total de execução da tarefa. Já a extensão CUDA é responsável por viabilizar a utilização do poder de processamento das placas gráficas, assim agilizando a execução do software. Com o pacote cuRAND somos capazes de utilizar os processadores das placas gráficas para gerar números aleatórios mais específicos.

Figura 2 – Comunicação dos computadores (programação paralela)



Fonte: Elaborado pelos autores (2023).

O Esquema da figura 2 representa um ciclo de iteração para simulação multicanônica paralela: N processadores (nós) são inicializados com diferentes sementes aleatórias e, assim, evoluem cadeias de Markov independentes, cujos histogramas individuais de contagem de energias configuracionais são acumulados em $H(E)$ a cada iteração MUCA para atualizar o peso $W(E)$. Os coeficientes resultantes (α, β) nos pesos MUCA são então computados no Mestre e transmitidos em broadcast, atualizando cada nó para que possa iniciar o próximo passo iterativo. Repete-se o processo até que $H(E)$ vs E torne-se plano e, portanto, a densidade de estados $\Omega(E)$ seja equiprovável para todas as Energias.

Fórmulas utilizadas

O software PHAST se baseia nas seguintes fórmulas para realizar suas simulações: entropia, curva calórica, energia livre e calor específico. Todas as fórmulas citadas estão em ordem logo abaixo.

$$S(E) = k_b \ln \Omega(E) \quad (1)$$

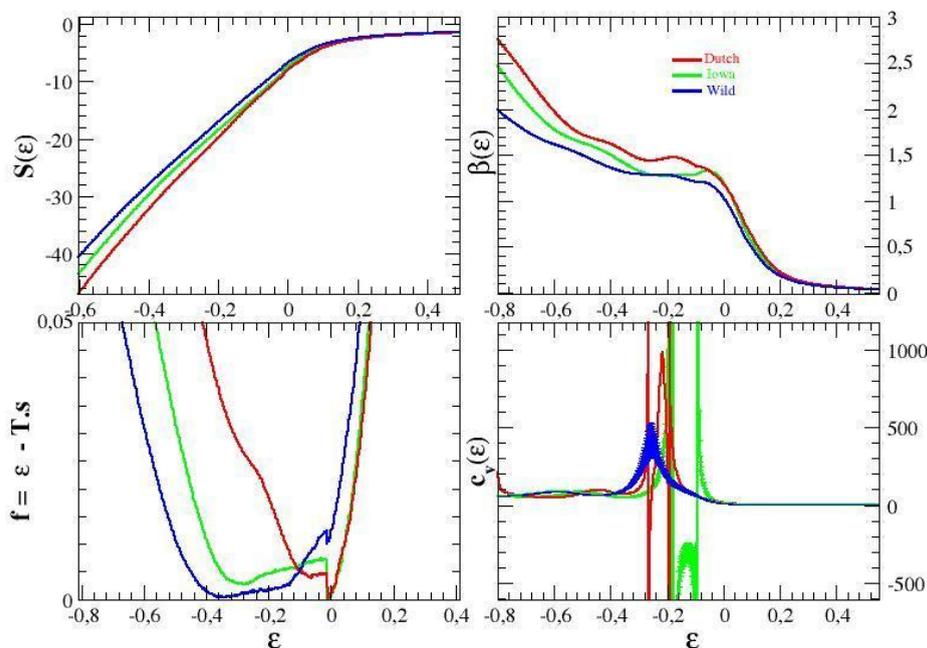
$$k_b \beta \equiv T(E)^{-1} = \frac{\partial S}{\partial E} \quad (2)$$

$$F(E) = E - \left(\frac{\partial S}{\partial E} \right)^{-1} \Big|_{E=E(T_c)} S(E) \quad (3)$$

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

RESULTADOS OBTIDOS

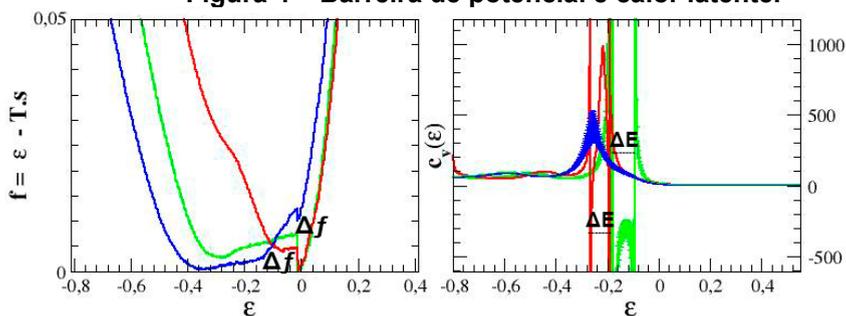
Figura 3 – Gráficos gerados pelo software PHAST



Fonte: Elaborado pelos autores (2023).

Os gráficos da figura 3 representam respectivamente: entropia $S(\epsilon)$, curva calórica $\beta(\epsilon)$, energia livre $f(\epsilon, T, s)$ e calor específico $C_v(\epsilon)$. A probabilidade de uma proteína agregar-se está diretamente relacionada a sua altura de barreira de potencial e ao seu calor latente, quanto maior a barreira de potencial e o calor latente menos propensa a agregação de proteína.

Figura 4 – Barreira de potencial e calor latente.



Fonte: Elaborado pelos autores (2023).

Ao analisar os gráficos da figura 4, percebemos que a proteína Wild-Type possui uma transição de fase de segunda ordem enquanto as mutações Dutch e Iowa possuem uma transição de fase de primeira ordem.

Nossos resultados são coerentes com resultados experimentais recentes, como *Illes-Toth et al (2021)* que monitoram a agregação de variantes $A\beta_{42}$ por espectrometria de massas. Estes autores constataram que, em ordem decrescente de agregação, temos as mutantes: Iowa (D23N), Italian (E22K) e Dutch (E22Q).

CONCLUSÃO

Os resultados obtidos na análise nos permitem concluir que, de acordo com a barreira de potencial (Δf) e calor latente (ΔE), a proteína mais propensa a se agregar é a Dutch seguida de Iowa. Isso significa que possivelmente organismos com maior quantidade de proteína beta-amilóide do tipo Dutch possuem maior probabilidade de serem acometidos pelo Alzheimer.

AGRADECIMENTOS

Agradeço a faculdade UTFPR campus Toledo-PR e ao CNPq pela oportunidade e apoio financeiro. Também agradeço o professor Rafael Frigori pelo grande auxílio e paciência durante a jornada de elaboração dessa iniciação científica.

DISPONIBILIDADE DE CÓDIGO

O software PHAST pode ser encontrado no seguinte link:
<https://data.mendeley.com/datasets/ggds2grzw9/1>.

CONFLITO DE INTERESSE

Não há conflito de interesse.

REFERÊNCIAS

CHAPMAN, S. J.. **Fortran for Scientists and Engineers** 4^o Edição. ISBN 9780073385891 (2017).

FRIGORI, R. B.. PHAST: **Protein-like heteropolymer analysis by statistical thermodynamics**, Computer Physics Communications. publicação 215 p.165-172. (2017).

ILLES-TOOTH, Eva et al. **Pulsed hydrogen–deuterium exchange reveals altered structures and mechanisms in the aggregation of familial Alzheimer’s disease mutants**. ACS Chem. Neurosci. 2021, 12, 1972–1982. DOI: 10.1021/acscchemneuro.1c00072. Disponível em: Pulsed Hydrogen–Deuterium Exchange Reveals Altered Structures and Mechanisms in the Aggregation of Familial Alzheimer’s Disease Mutants | ACS Chemical Neuroscience. Acesso em: 16 ago.2023.

SILVA, P. G., BIANCHINI, C. P., COSTA, E. B.. **Programação Paralela e Distribuída: com MPI, OpenMP e OpenACC para computação de alto desempenho**. ISBN 9788555193033 (2022).