



SEI-SICITE 2021

Pesquisa e Extensão para um mundo em transformação

Simulações Multi Canônicas de Proteínas

MULTI-CANONICAL PROTEIN SIMULATIONS

Fernanda Azevedo*, Rafael Bertolini Frigori†

RESUMO

As simulações *in silico* auxiliam no estudo termo-estrutural das proteínas a fim de se obter mais informações sobre o SARS-Cov-2, responsável pelo Covid-19 (doença respiratória grave que pode ser transmitida através do ar ou contato), foi utilizado o pacote PHAST que trata-se de um pacote de código aberto que permite que o usuário simule e realize a análise termo-estatísticas micro canônicas de polímeros lineares semi-flexíveis e proteínas (via modelagem de granulação grossa), para realizar simulações de Monte Carlo a fim de realizar a simulação e a análise dos dados obtidos. O objetivo do trabalho foi realizar simulações e a análise dos dados utilizando os módulos PHAST e ANALYST do software, através da análise dos gráficos podemos concluir que o RBM da variante SARS2/WT é termodinamicamente menos estável (i.e. seu calor latente de enovelamento, ou L , é menor) do que o da “variante inglesa” B.1.1.7(α) e portanto é estruturalmente mais flexível. Dessa forma observamos que a variante α tem maior estabilidade e potencial de evadir anticorpos, portanto pior impactando o cenário pandêmico, como visto na Inglaterra, dada a sua maior infectividade.

Palavras-chave: Simulação, Proteína, SARS-Cov-2, Física Estatística.

ABSTRACT

The *in silico* simulations help in the thermo-structural study the protein variations in order to get more information about SARS-Cov-2, responsible for Covid-19 (severe respiratory disease that can be transmitted by air or contact), the PHAST package was used, which is an open source package that allows the user to simulate and perform micro canonical thermo-statistical analysis of semi-flexible linear polymers and proteins (via coarse-grained modelling), to perform Monte Carlo simulations in order to perform the simulation and analysis of the data obtained, the objective of the work was to perform simulations and data analysis using the software's PHAST and ANALYST modules, through the analysis of the graphics we can conclude that the RBM of the SARS2/WT variant it is thermodynamically less stable (i.e. its latent heat of folding, or L , is lower) than that of B.1.1.7(α) and therefore it is structurally more flexible. Thus, we observe that the α variant has greater stability and the potential to evade antibodies, therefore having a worse impact on the pandemic scenario, as seen in England, given its greater infectivity.

Keywords: Simulation, Protein, SARS-Cov-2, Statistical Physics.

1. INTRODUÇÃO

Os polipeptídeos são cadeias lineares hetero-poliméricas de aminoácidos, enquanto as proteínas são uma grande classe de polímeros biológicos, contendo pelo menos um polipeptídeo que juntos formam um conjunto de macromoléculas e exercem várias funções dentro do organismo. A hipótese termodinâmica afirma que a estrutura original da proteína é única e estável cineticamente, sua energia livre é determinada por sua sequência primária de aminoácidos (BRANDEN; TOOZE, 1998).

O mal-dobramento das proteínas pode produzir partes disfuncionais que induzem a formação de amilóides citotóxicos que podem culminar em doenças degenerativas como o Diabetes Tipo II e o Alzheimer. Como os biopolímeros são pequenos sistemas podem ser realizados cálculos de densidades de estados a fim de realizar investigações sobre esses sistemas termodinâmicos no ensemble microcanônico. Utilizar tal técnica implica em um grande acúmulo de dados estatísticos por meio dos métodos de Monte Carlo como Wang-Landau ou

* Engenharia de Computação, Universidade Tecnológica Federal do Paraná, Toledo, Paraná, Brasil; fazevedo@alunos.utfpr.edu.br

† Universidade Tecnológica Federal do Paraná, Campus Toledo; frigori@utfpr.edu.br



o ensemble multi canônico (MUCA), tais problemas podem ser resolvidos através de eficientes algoritmos paralelos e granulação grossa dos campos de força molecular (LANDAU; BINDER, 2005).

As simulações *in silico* também podem auxiliar no estudo de variantes de uma mesma proteína, a fim de obter-se um maior entendimento de como e porquê as mutações/variantes surgiram, como é o caso da proteína S do SARS-COV-2 (responsável pela invasão viral no interior celular) que após 2019 apresentou diversas mutações em vários locais do mundo (YI; *et al*, 2020 e HAUSER; *et al*, 2021 e PREVENTION,2021).

O Pacote PHAST (FRIGORI, 2017) foi desenvolvido com o objetivo de permitir massivas simulações paralelas de Monte Carlo, com respectivo acúmulo de grandes bancos de dados, trata-se de um pacote de código aberto que permite que o usuário simule e realize a análise termo-estatísticas micro canônicas de polímeros lineares semi-flexíveis e proteínas (via modelagem de granulação grossa).

O objetivo deste trabalho de iniciação científica é realizar simulações *in silico* de variantes da proteína Spike do SARS-Cov-2 (Covid-19) via simulações de Monte Carlo e sua posterior análise e interpretação micro canônica. Empregamos os módulos PHAST (simulação) e ANALYST (análise) do software para tais tarefas.

2.PROCEDIMENTOS OPERACIONAIS DA PESQUISA

Simulação

As simulações de Monte Carlo são importantes não apenas para caracterizar transições de fases estruturais mas também distinguir entre “bons” e “maus” enoveladores, a classificação depende de uma determinação precisa da densidade de estados Ω dos sistemas que pode ser encontrada utilizando Wang-Landau ou com ensembles generalizados (MUCA) (LANDAU; BINDER, 2005 e FRIGORI, 2017). Uma vez que Ω passa a ser conhecida, como função da energia interna E, a termodinâmica micro canônica pode ser estabelecida pela fórmula usual de Boltzmann para entropia S(E):

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

Dessa forma as derivadas numéricas da entropia podem ser utilizadas para calcular quantidades de interesse como a curva calórica micro canônica, relacionando a temperatura T, a energia E interna:

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

Também pode-se obter o calor específico $C_V(E)$ e a energia livre F(E):

$$C_V(E) = \frac{dE}{dT} = - \left(\frac{\partial S}{\partial E} \right)^2 \left(\frac{\partial^2 S}{\partial E^2} \right)^{-1} \quad (3)$$

$$F(E) = E - \left(\frac{\partial S}{\partial E} \right)^{-1} \Big|_{E=E(T_c)} S(E) \quad (4)$$

Onde T_c é a temperatura pseudo-crítica (fixa) de transição de fase.

A entropia, como pode-se observar, é escrita como uma função por partes das energias discretizadas E_k e um conjunto de coeficientes MUCA $\{\beta_k, \alpha_k\}$:

$$S_{muca}(E_k) = \beta_k E_k - \alpha_k \quad (5)$$

Os coeficientes MUCA aumentam o tunelamento energético através de barreiras de potencial termodinâmico (presentes em transições de fase), melhorando assim a amostragem configuracional. A determinação de tais coeficientes exige que um algoritmo iterativo que envolve o cômputo de histogramas de energia, para isso no início ajusta-se para um valor inicial unitário todas as energias E_k que utilizadas para uma simulação, depois uma recursão atualizada os pesos MUCA antes de se começar uma nova simulação.



O PHAST emprega histogramas ponderados por erros, então as estatísticas para calcular os pesos melhoram a cada repetição, enquanto a convergência é alcançada pela recursão de peso de Berg. O estimador paralelo foi proposto para pesos MUCA (ZIERENBERG, 2014), para estudar sistemas de spin que apresentam propriedades que excedem as abordagens mais tradicionais. Esse algoritmo foi implementado por Frigori (2017) no pacote PHAST, assim, o problema pode ser solucionado com um esquema de atualização simples, onde primeiramente cada processador é inicializado com os mesmos coeficientes MUCA mas uma semente aleatória particular, então de maneira independente eles executam a simulação MCMC (Markov Chain Monte Carlo) (LANDAU; BINDER, 2005 e FRIGORI, 2017) para acumular histogramas de energia local, esses serão combinados no nó mestre que atualiza os novos pesos e repete o ciclo até que a convergência seja alcançada.

Software

O pacote utilizado é constituído dos seguintes módulos: SET_INPUT, PHAST e ANALYST (FRIGORI, 2017):

- SET_INPUT: é responsável por preparar arquivos de entradas das proteínas.
- PHAST: é responsável por realizar as simulações paralelas.
- ANALYST: é responsável por calcular as termo-estatísticas para as simulações que foram feitas anteriormente pelo MUCA, quando o ANALYST é chamado, o arquivo last.con é lido e fornece ao software os parâmetros que foram definidos em vars.h, usado na simulação anterior, dessa forma o usuário tem a opção de escolher os intervalos de configuração para os arquivos a serem analisados.

SARS-Cov-2

O SARS-Cov-2 é um vírus de RNA responsável pela Covid-19, uma grave doença respiratória, que pode ser transmitida pelo ar ou através do contato. Em dezembro de 2019 a doença foi identificada em humanos em Wuhan, China, e logo tornou-se pandêmica. Após a molécula Spike do SARS-Cov-2 ter sido sequenciada em laboratório, os microbiologistas concluíram que o SARS-Cov-2 está filogeneticamente classificado no mesmo gênero SARS, do qual o SARS-Cov-1 (identificado em 2002, na China) também pertence. As variantes de interesse requerem uma atenção maior, e uma caracterização laboratorial mais aprimorada, dada sua maior evasão imune e contagiosidade. A classificação e a análise dessas variantes é importante para o desenvolvimento e testes de novas vacinas, diagnóstico molecular de doenças e profilaxias (YI; *et al*, 2020 e HAUSER; *et al*, 2021 e PREVENTION,2021).

Segundo Yi, *et al.*(2020) estudos mostraram que o SARS-Cov-2 utiliza a glicoproteína de pico homotrimérico (S) ou proteína spike viral para se ligar ao receptor funcional ACE2 (hACE 2) humano, e portanto acredita-se que esse seja o principal alvo para bloquear a entrada viral, o domínio estrutural RBD na proteína S faz a intermediação entre a ligação do vírus às células hospedeiras, de acordo com estudos sobre a estrutura cristalina, o *motif* de ligação do receptor (RBM) é o principal *motif* funcional em RBD sendo composto por duas regiões que formam a interface entre a proteína S e hACE 2, e embora a identidade dos aminoácidos na região RBM seja baixa, o mecanismo de ligação é semelhante nos vírus SARS-Cov-1 e SARS-Cov-2.

3.RESULTADOS

No desenvolvimento do trabalho foram realizadas com o Pacote PHAST simulações *in silico* do RBM da proteína S de variantes de preocupação do SARS-Cov-2, a saber: Alfa e WT, que podem ser visualizadas abaixo.

- RBM Spike Protein-SARS2 - B.1.17(α):

Antes do mapeamento:

SNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPT**Y**GVGYQPY

Após o mapeamento pela escala Roseman pelo SET_INPUT:

BBBABBABBBBBBABBAABBBBABAABBBABABABBABBAAABBABBABABAAABBBBABAA**A**BABBBA
B

- RBM Spike Protein-SARS2 - WT:

Antes do mapeamento:

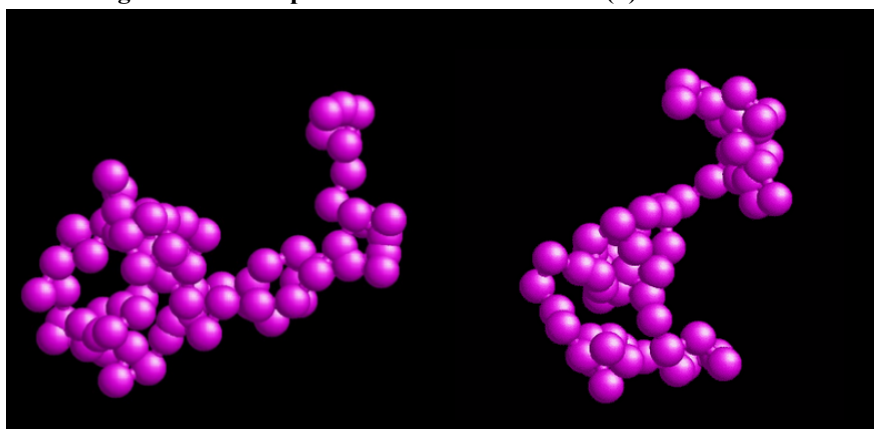
SNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPT**N**GVGYQPY

Após o mapeamento pela escala Roseman pelo SET_INPUT:

BBBABBABBBBBBABBAABBBBABAABBBABABABBABBAAABBABBABABAAABBBBABAA**B**BABBBA
B

Dessa maneira o PHAST foi utilizado para realizar as simulações em paralelo das proteínas, para o presente trabalho utilizamos até 701 iterações multicanônicas de ambas as proteínas. Após isso as configurações foram analisadas pelo ANALYST com ambas utilizando os pesos multicanônicos produzidos entre as iterações 551 e 701. Dessa forma a análise da variante B.1.17 utilizou os parâmetros $dE=1.5$ e $\beta_{\text{crítico}}=0.97$, valores que foram encontrados após análises utilizando o xmgrace, enquanto isso para WT foram utilizados os seguintes valores: $dE=1.5$ $\beta_{\text{crítico}}=1.21$. Na Figura 1 produzida com o Rasmol 2.7.5.2 podemos observar as estruturas nativas das moléculas RBM Spike Protein B.1.17(α) (a esquerda) e WT (a direita) respectivamente:

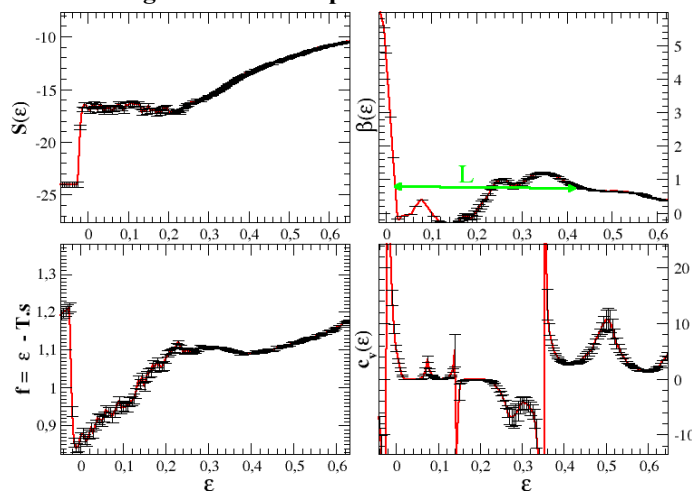
Figura 1- RBM Spike Protein-SARS2/B.1.1.7(α) e SARS2/WT



Fonte: do Autor (2021)

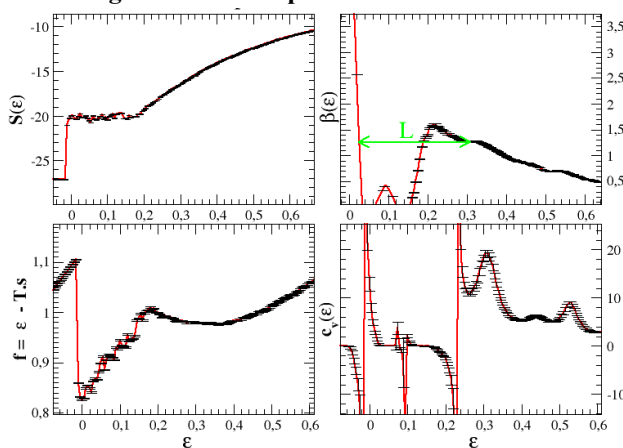
Os demais parâmetros utilizados na análise foram encontrados após análises de ajuste fino para obtermos os resultados mais precisos em relação ao $\beta_{\text{crítico}}$ e de forma que não houvesse interferência de fatores como a temperatura. Após a análise obtemos os gráficos apresentados nas figuras abaixo.

Figura 2 - RBM Spike Protein-SARS2/alfa



Fonte: do Autor (2021)

Figura 3 - RBM Spike Protein-SARS2/WT



Fonte: do Autor (2021)

Podemos observar nos gráficos acima gerados pelo ANALYST os valores de entropia $S(\epsilon)$, a curva calórica $\beta(\epsilon)$, a energia livre deslocada $f = \epsilon - T.s$. (à temperatura crítica $\beta_{\text{critic}}^{-1}$) e o calor específico $c_v = L$.

Observa-se na Figura 2 que as transições de fase do sistema são de primeira ordem (curva calórica possui um “S-bend”) e a variante possui um calor específico de aproximadamente 0,372. Já na Figura 3 temos que as transições de fase são de primeira ordem mas o calor específico é de aproximadamente 0,262.

Através da análise dos gráficos concluímos que o RBM da variante SARS2/WT é termodinamicamente menos estável (i.e. seu calor latente de envelhecimento, ou L , é menor) do que o da B.1.1.7(α) e portanto é estruturalmente mais flexível. Desse modo a variante α tem maior estabilidade e potencial de evadir anticorpos, portanto podendo pior impactar o cenário pandêmico, como visto na Inglaterra, dada a sua maior infectividade (associada à maior estabilidade molecular).

4.CONCLUSÃO

Neste trabalho de iniciação científica realizamos simulações in silico de variantes da proteína Spike do SARS-Cov-2 (Covid-19) via simulações de Monte Carlo e uma posterior análise e interpretação



termodinâmica microcanônica, empregando os módulos PHAST e ANALYST, além disso mostramos as transições de fase e calor específico (microcanônico) do sistema, e através de análises podemos concluir qual variante de preocupação, ou VOC, é mais transmissível (α), aplicando conceitos de Física Estatística estudados ao decorrer deste projeto.

AGRADECIMENTOS

Gostaria de agradecer a Universidade Tecnológica Federal do Paraná pelo apoio financeiro e a orientação realizada pelo Prof. Dr. Rafael B. Frigori durante todo o processo, e ao apoio da minha família.

REFERÊNCIAS

- BRANDEN, Carl Ivar; TOOZE, John. **Introduction to Protein Structure**. 2. ed. New York: Garland Science, 1998. 424 p. (9780815323051).
- CAGLIANI, Rachele; *et al.* **Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2**. Journal Of Virology, [S.L.], v. 94, n. 12, p. 0-100, 01 jun. 2020. American Society for Microbiology. <http://dx.doi.org/10.1128/jvi.00411-20>
- FRIGORI, Rafael B.. PHAST: protein-like heteropolymer analysis by statistical thermodynamics. **Computer Physics Communications**, [S.L.], v. 215, p. 165-172, jun. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.cpc.2017.01.021>
- HAUSER, Blake M; *et al.* **Rationally designed immunogens enable immune focusing to the SARS-CoV-2 receptor binding motif**. bioRxiv 2021.03.15.435440; doi: <https://doi.org/10.1101/2021.03.15.435440>
- LAN, Jun; *et al.* **Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor**. **Nature**, [S.L.], v. 581, n. 7807, p. 215-220, 30 mar. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41586-020-2180-5>.
- LANDAU, David P.; BINDER, Kurt. **A Guide To Monte Carlo Simulation Statistic Physics**. 2. ed. New York: Cambridge University Press, 2005. (13978-0-511-13098-4). Disponível em: <https://dokumen.tips/documents/a-guide-to-monte-carlo-simulations-in-statistical-physics.html>. Acesso em: 20 fev. 2021.
- MEINKE, J., MOHANTY, S., EISENMENGER, F., HANSMANN, U. **SMMP v. 3.0 Simulating proteins and protein interactions in Python and Fortran**. **Computer Physics Communications**, 178 (6) (2008), pp. 459-470
- PREVENTION, Centers For Disease Control And. **SARS-CoV-2 Variant Classifications and Definitions**. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>. Acesso em: 01 jul. 2021.
- SHANG, Jian; *et al.* **Structural basis of receptor recognition by SARS-CoV-2**. **Nature**, [S.L.], v. 581, n. 7807, p. 221-224, 30 mar. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41586-020-2179-y>.
- ZIERENBERG, Johannes; MARENZ, Martin; JANKE, Wolfhard. **Scaling Properties of Parallelized Multicanonical Simulations**. **Physics Procedia**, [S.L.], v. 53, p. 55-59, 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.phpro.2014.06.026>.
- YI, Chunyan; *et al.* **Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies**. **Cellular & Molecular Immunology**, [S.L.], v. 17, n. 6, p. 621-630, 15 maio 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41423-020-0458-z>.