



Dinâmica Molecular no estudo de Proteínas

Molecular Dynamics in the study of Proteins

Nívea Neres Ferreira de Jesus, Rafael Bertolini Frigori

RESUMO

A dinâmica molecular é um método que descreve o comportamento dos átomos e moléculas em função do tempo, a simulação desse processo no presente trabalho foi realizada com auxílio do pacote GROMACS, onde analisamos três variações do segmento RBM encontrado na proteína spike do vírus SARS-Cov-2. Para preparação do sistema, centralizamos a proteína em uma caixa e definimos suas dimensões com 2,0nm entre as imagens periódicas da proteína e a tratamos como célula unitária. Em seguida realizamos a adição de íons e a minimização de energia garantindo que não houvesse choque estéricos no sistema, passando para as fases de equilíbrio que se dividiu em duas etapas, do conjunto de Número constante de partículas, Volume e Temperatura (NVT) e do conjunto de Número de partículas, Pressão e Temperatura (NPT). Após finalizar as fases de equilíbrio, as restrições de posição foram liberadas e a produção da dinâmica molecular executada para coleta de dados para análise, onde concluímos que existem diferenças de estabilidade, equilíbrio e flexibilidade entre os segmentos RBM. Nossos resultados são condizentes com recentes descobertas empíricas.

Palavras-chave: GROMACS, RBM, proteína, spike.

ABSTRACT

Molecular dynamics is a method that describes the behavior of atoms and molecules as a function of time. The simulation of this process in the present work was performed with the aid of the GROMACS package, where we analyzed three variations of the RBM segment found in the spike protein of the SARS-Cov virus -two. To prepare the system, we centered the protein in a box and defined its dimensions with 2.0nm between the periodic images of the protein and treated it as a unit cell. Then, we performed the addition of ions and the minimization of energy, ensuring that there was no steric shock in the system, moving to the equilibrium phases that were divided into two stages, the set of Constant Number of Particles, Volume and Temperature (NVT) and the Particle Number, Pressure and Temperature (NPT) set. After completing the equilibrium phases, the position constraints were released and the production of molecular dynamics performed for data collection for analysis, where we concluded that there are differences in stability, equilibrium and flexibility between the RBM segments. Our results agree with recent empirical findings.

Keywords: GROMACS, RBM, protein, spike.

1 INTRODUÇÃO

A doença COVID-19 causada pelo vírus SARS-Cov-2, possui uma alta taxa de transmissão e se disseminou rapidamente de forma global, a replicação viral causa mutações genéticas, gerando assim inúmeras variantes. No final de 2020 e início de 2021 três novas linhagens de SARS-CoV-2 foram identificadas mundialmente e consideradas Variant of Concern (VOC). A variante B.1.1.7 foi identificada no Reino Unido e carrega a mutação proteica N501Y que aumenta a afinidade do vírus pelo receptor ACE-2. A variante B.1.351 foi identificada na África do Sul, e carrega três importantes mutações, parecendo ser mais transmissível e menos vulnerável a anticorpos gerados por uma infecção anterior ou por vacina. E a variante P.1 identificada



pela primeira vez em quatro viajantes no Japão, que retornavam do estado do Amazonas (Brasil), em 2 de janeiro de 2021 (FREITAS, 2021).

Por sua vez, a biologia molecular afirma que 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). Neste trabalho em particular, estudamos um pequeno segmento da proteína Spike do vírus SARS2, conhecido como RBM, esse segmento é o principal responsável pela intrusão viral no interior celular. É contra ele que agem os chamados anticorpos neutralizantes, formados por vacinas, sendo suas mutações as principais responsáveis pela “evasão viral” (i.e. falha vacinal). Pela falta de estruturas cristalográficas no *Protein Data Bank*, modelamos essas estruturas proteicas do vírus original (WT) e suas variantes B.1.1.7, B.1.351-P1 por homologia, usando o software I-TASSER (YANG, 2015).

Até o início desse projeto, não haviam simulações de Dinâmica Molecular de RBMs de VOCs como referência para nossos estudos. Vale destacar que, o processo de simulação computacional foi realizado através do pacote GROMACS, desenvolvido para realizar simulações de dinâmica molecular e minimização de energia. Duas técnicas que pertencem ao reino da química computacional e modelagem molecular. Essas técnicas permitem prever com precisão os movimentos de átomos e moléculas através das equações de Newton. As Eq. (1) e Eq. (2) serão resolvidas simultaneamente durante intervalos de tempo limitados, para garantir que os valores de temperatura e pressão sejam adequados, essas informações ficam registradas em um arquivo de saída com as coordenadas em função do tempo, definindo a trajetória do sistema. Esta trajetória alcançará um estado de equilíbrio, onde podemos extrair muitas propriedades macroscópicas (ABRAHAM, 2016).

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, \quad i = 1 \dots N \quad (1)$$

$$F_i = - \frac{\partial V}{\partial r_i} \quad (2)$$

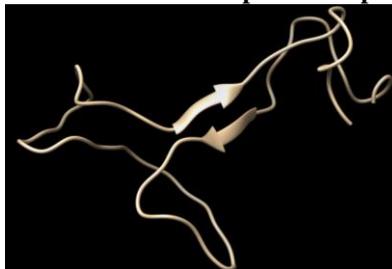
Na minimização de energia, o GROMACS faz uso das informações de derivadas parciais emitidas durante a dinâmica molecular para encontrar um conjunto de coordenadas que minimizam a energia potencial do sistema, percorrendo a superfície de potencial na direção em que a energia diminui, fazendo com que o sistema seja elevado a um mínimo de energia local através de mudança nas posições atômicas, o que ocasiona em um relaxamento entre os ângulos e nas distorções das ligações químicas. Essa técnica é muito importante para a orientação da geometria e do solvente, garantindo que a simulação não se desenvolva a partir de um ponto desestabilizado energeticamente.

2 MÉTODO

Para realizar as simulações dos sistemas com as variantes dos segmentos RBM, adaptamos e seguimos o roteiro simulacional de LEMKUL (2021), contudo escolhemos como campo de força o Charmm22 e TIP4P/Ew como solvente explícito, pois são o estado-de-arte para simulação de pequenos peptídeos. Os comandos descritos a seguir para uso do GROMACS são para o modelo WT, porém foram adaptados e realizados de maneira análoga para os demais sistemas.

O arquivo com a estrutura da proteína, ilustrada na figura 1, é submetido ao GROMACS usando o comando `gmx pdb2gmx -ignh -f WT_modell.pdb -o WT_modell.gro`, gerando a topologia da molécula, um documento de restrição de posição e da estrutura pós-processada.

Figura 1 - Estrutura do RBM da proteína Spike do SARS-2



Fonte: Autoria própria.

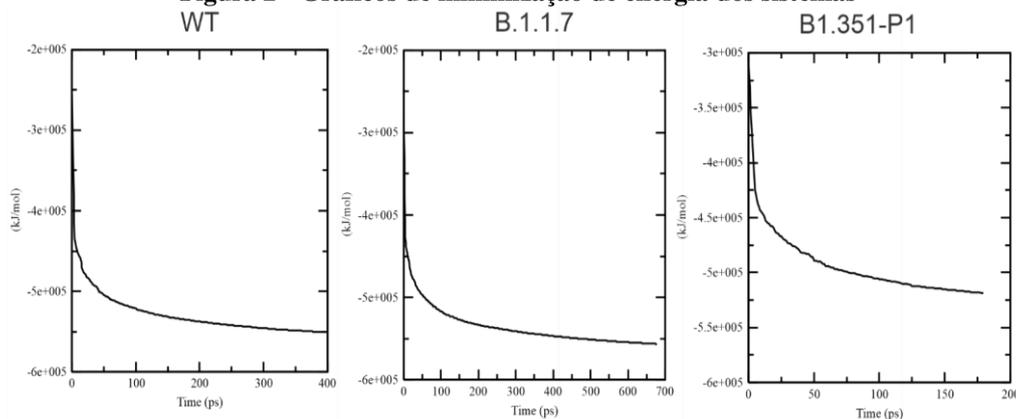
Em seguida a proteína foi centralizada em uma caixa tratada como célula unitária, usando o comando `gmx editconf -f WT_modell.gro -o WT_newbox.gro -c -d 1.0 -bt cubic` e definimos suas dimensões com 2,0nm entre suas imagens periódicas e a enchemos com solvente, com o comando `gmx solvate -cp WT_newbox.gro -cs tip4p.gro -o WT_solv.gro -p topol.top`, onde o solvato controlou as mudanças feitas quando as moléculas foram inseridas, deixando o sistema com uma carga líquida de +1e na WT e B.1.1.7, +3e na B.1.351-P1, onde adicionamos o número de íon de Cl⁻ correspondentes a cada uma para balancear o sistema através do genion com o comando `gmx genion -s ions.tpr -o WT_solv_ions.gro -p topol.top -pname NA -nname CL -neutral`.

Para garantir que não acontecesse choques estéricos ou deixasse a geometria inadequada foi realizado o processo de minimização de energia através do comando `gmx mdrun -v -deffnm em`, gerando arquivos de energia e trajetória binária, arquivos da estrutura com a energia minimizada e de registro de processo. Depois de assegurar que a energia potencial e a força máxima estivessem dentro dos valores adequados para prosseguir com a simulação, realizamos uma análise utilizando comando `gmx energy -f em.edr -o potential.xvg`, que gerou os gráficos de minimização de energia (Figura 2).

Realizada a minimização de energia precisamos garantir o equilíbrio do solvente e dos íons em torno da proteína. Esse equilíbrio é realizado em duas fases, a primeira é o conjunto do Número constante de partículas, Volume e Temperatura - NVT, que utiliza o arquivo `pose.itp` gerado no primeiro processo da simulação, para aplicar uma força restritiva de posição sobre os átomos pesados da proteína e equilibrar o sistema. Utilizando novamente o `grompp` e o `mdrun` realizamos a análise da temperatura do sistema, através dos comando `gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr` e `gmx mdrun -deffnm nvt`, gerando o gráfico do equilíbrio do NVT (Figura 3).

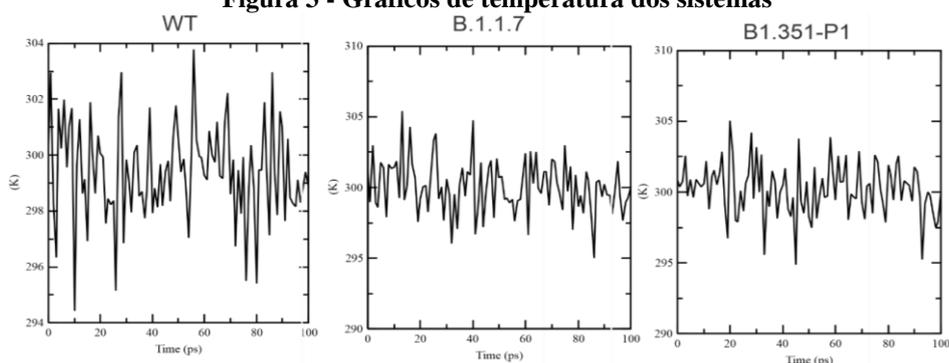
A segunda fase de equilíbrio é definida pelo número constante de partículas, pressão e temperatura - NPT, onde conduzimos um processo isobárico utilizando os comandos `gmx grompp -f npt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt.tpr` e `gmx mdrun -deffnm npt`, e analisamos a progressão da pressão através dos gráficos na (Figura 4).

Figura 2 - Gráficos de minimização de energia dos sistemas



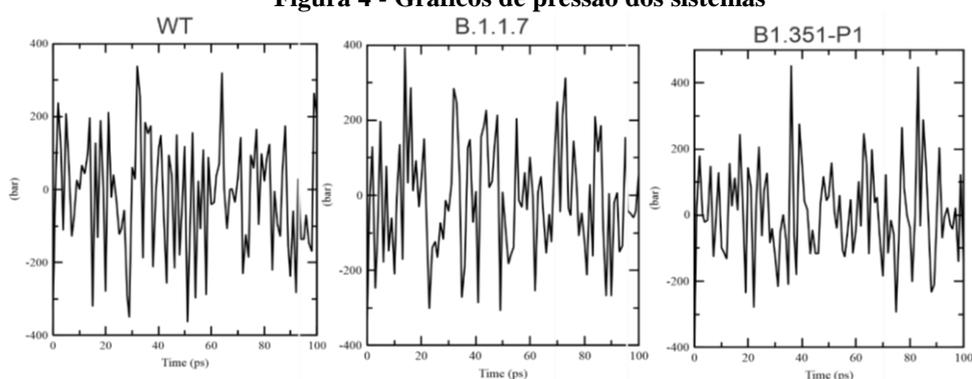
Fonte: Autoria própria.

Figura 3 - Gráficos de temperatura dos sistemas



Fonte: Autoria própria.

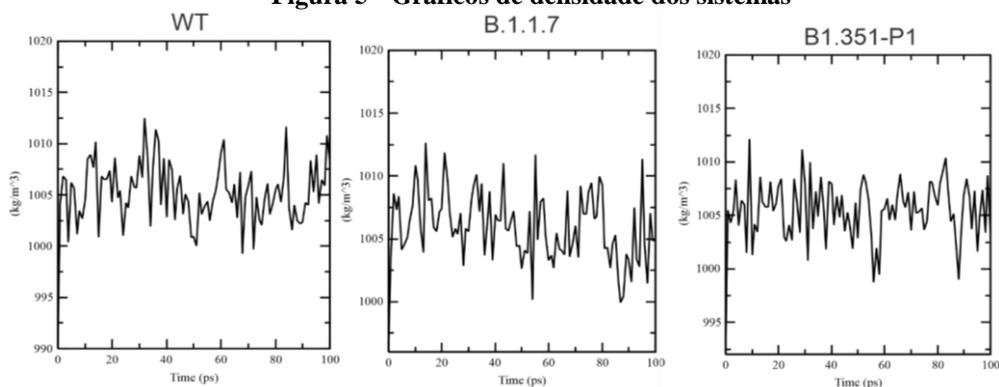
Figura 4 - Gráficos de pressão dos sistemas



Fonte: Autoria própria.

Os valores de pressão variam bastante durante os 100ps da fase de equilíbrio, o que dificulta apontar uma diferença entre os valores obtidos. Porém o fim deste processo também torna possível verificar a variação da densidade no processo através dos gráficos gerados por *gmx energy -f npt.edr -o density.xvg*, (Figura 5)

Figura 5 - Gráficos de densidade dos sistemas



Fonte: Autoria própria.

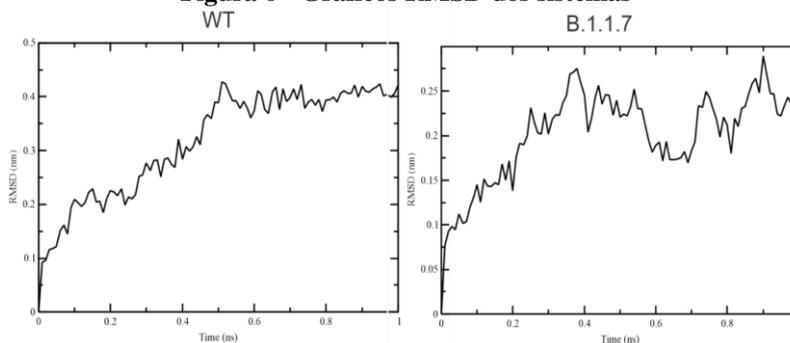
Com o sistema bem equilibrado nos aspectos de temperatura e pressão, precisamos liberar a restrição de posição feita anteriormente e realizar por fim o processo de dinâmica molecular com *gmx mdrun -deffnm md_0_1*, obtendo os dados necessários para realizar a análise.



3 RESULTADOS

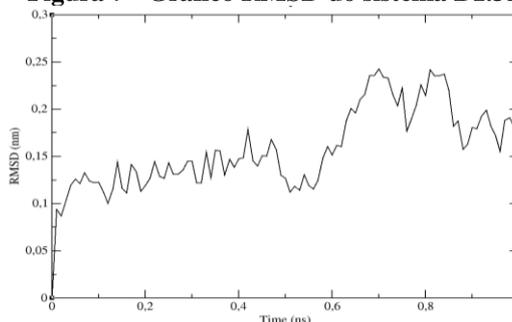
Para verificação e análise dos resultados obtidos, um fator muito importante é o valor médio de desvio dos átomos da estrutura (RMSD), pois este valor simboliza as diferenças estruturais das proteínas como estabilidade e equilíbrio do sistema. Podemos visualizar nos gráficos a seguir (Figuras 6 e 7) o desempenho das proteínas simuladas em relação a essa propriedade.

Figura 6 - Gráficos RMSD dos sistemas



Fonte: Autoria própria.

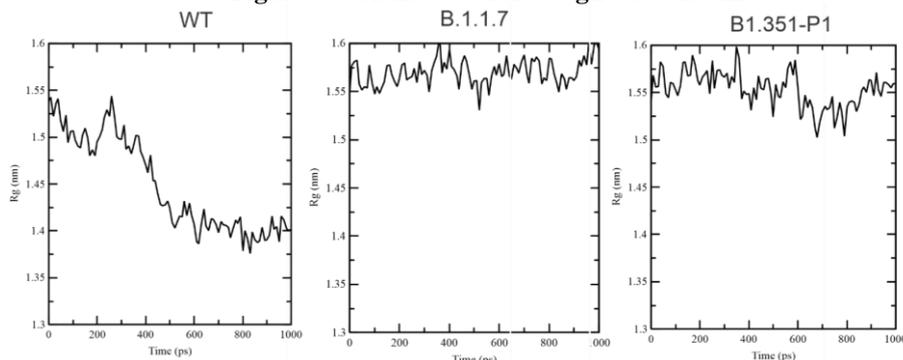
Figura 7 - Gráfico RMSD do sistema B1.351-P1



Fonte: Autoria própria.

Outra propriedade relevante obtida é o raio de giro, pois é a medida do nível de compactação e descreve a evolução da geometria da proteína durante o período de simulação, informações que podem ser visualizadas a seguir na (Figura 8).

Figura 8 - Gráficos de raio de giro dos sistemas



Fonte: Autoria própria.



4 CONCLUSÃO

Através das simulações pudemos concluir que no aspecto de estabilidade termodinamicamente e equilíbrio estrutural a variante B1.351-P1 é mais estável, pois seus níveis de RMSD permanecem entre 0,05nm a 0,25nm durante o período de 1ns, sendo dos três sistemas o mais próximo a zero. E ao mesmo tempo mais flexível estruturalmente, visto nos gráficos de raio de giro dos sistemas é possível perceber que as variações de RBM B1.351-P1 e WT apresentam as maiores oscilações, enquanto a proteína B.1.1.7 tem menor variação durante o processo de simulação. Esses fatores são importantes para permitir maior habilidade de ligação com a hACE2, o que tornaria as variantes B1.351 e P1 as mais contagiosas. Os resultados computacionais observados em nossos estudos são ainda qualitativamente compatíveis com estudos empíricos, recentes, que empregaram técnicas de mutações genéticas induzidas laboratorialmente (STARR, 2021). Isso é particularmente encorajador pois demonstra que a Dinâmica Molecular é uma técnica acurada e que poderá ser útil, eventualmente auxiliando, no futuro desenho racional *in silico* de anticorpos monoclonais para terapias mais efetivas contra a Covid-19.

AGRADECIMENTOS

Gostaria de agradecer o apoio da Fundação Araucária FA - Paraná/Brasil para realização deste trabalho e ao professor Rafael Bertolini Frigori por todo o apoio durante o projeto.

REFERÊNCIAS

- Lemkul, Justin A. **Gromacs Tutorial Lysozyme in Water**. Disponível em: <http://www.md-tutorials.com/gmx/lysozyme/>. Acesso em: 24 de julho de 2021
- BRANDEN, C. I.; TOOZE, J. **Introduction to protein structure**. New York: Garland Science, 1998.
- J Yang, Y Zhang. I-TASSER server: **new development for protein structure and function predictions**. Nucleic Acids Research, 43: W174-W181 (2015).
- Abraham, M, *et al*. Gromacs: Reference Manual. Disponível em: <https://manual.gromacs.org/>. Acesso em: 24 de julho de 2021
- Tracking of Variants - hCov19. GISAIID, 2021. Disponível em: <https://www.gisaid.org/hcov19-variants/>
- Freitas, A. R. R, *et al* . A emergência da nova variante P.1 do SARS-Cov-2 no Amazonas (Brasil), foi temporalmente associada a uma mudança no perfil da mortalidade devido a COVID-19, segundo sexo e idade. Disponível em: <https://doi.org/10.1590/SciELOPreprints.2030> . Acesso em: 6 de outubro de 2021.
- Starr, T. N, *et al*. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0092867420310035> . Acesso em: 7 de outubro de 2021.