

## CAUSAS TEÓRICAS DAS LINHAGENS BRASILEIRAS P.1 e P.2 DO VÍRUS SARS-COV-2 MEDIANTE DINÂMICA MOLECULAR

Micael D. L. Oliveira<sup>1,\*</sup>, Kelson M. T. Oliveira<sup>2</sup>, Jonathas N. Silva<sup>3</sup>, Clarice Santos<sup>4</sup>, João Bessa<sup>5</sup> e Rosiane de Freitas<sup>6</sup>

1. Estudante de IC do curso de Física - Universidade Federal do Amazonas (UFAM)
2. Professor/Pesquisador do Laboratório de Química Teórica e Computacional - UFAM / Orientador
3. Pesquisador do Laboratório de Química Teórica e Computacional - UFAM
4. Mestranda em Ciência da Computação - UFAM
5. Estudante de IC do curso de Ciência da Computação - UFAM
6. Professora/Pesquisadora do Laboratório de Otimização e Complexidade Computacional - UFAM

### Resumo

O novo  $\beta$ -coronavírus vem causando tristes perdas em todo o mundo e o surgimento de novas variantes é de grande preocupação. Assim, ao longo desta pesquisa foi estudada a linhagem B.1.1.28 de clado P.1 como também a P.2, onde através de dinâmica molecular no algoritmo NAMD3 no intervalo de 18 ns foi possível entender os impactos termodinâmicos na interação do complexo ACE2-RBD. Pela dinâmica molecular, notou-se que as médias de RMSF e RMSD na variante P.1 foram menos expressivas em comparação com a estrutura ausente de mutações, onde além disso, houve um aumento no número médio de ligações de Hidrogênio. Por meio da decomposição de energia MM/PBSA, descobriu-se também que as interações de Van der Waals predominaram e foram mais favoráveis quando a estrutura possui mutações da linhagem P.1. Portanto, a maior estabilização do complexo ACE2-RBD talvez seja uma explicação plausível do porquê algumas mutações estão convergindo em linhagens diferentes, como E484K e N501Y.

**Palavras-chave:** SARS-CoV-2; Dinâmica molecular; Linhagem P.1/P.2.

**Apoio financeiro:** CNPq.

### Introdução

Em janeiro de 2021, pesquisadores brasileiros identificaram uma nova variante do coronavírus, denominada P.1, durante um período de aumento sem precedentes nas internações e incidência de casos de reinfecção na cidade de Manaus, onde foi demonstrado estar correlacionado com o surgimento desta nova linhagem. A partir do sequenciamento genético percebeu-se consistir em múltiplas mutações, dentre as mais preocupantes estão 3 (três) que afetaram a região RBD da glicoproteína Spike: K417T, E484K e N501Y [1]. Além disso, percebeu-se ter havido um aumento da transmissibilidade em 1,4-2,2 vezes, além da maior probabilidade de letalidade em 1,1-1,8 vezes [2] além de um deslocamento para faixas etárias mais jovens.

Devido à alta frequência de determinadas mutações e as fortes pressões seletivas que afetam a evolução do coronavírus, é natural que mutações convergentes surjam, como por exemplo E484K. Desta forma há uma pressão seletiva regendo a evolução viral, que vai muito além de uma simples mutação aleatória [3].

Nesta pesquisa estudamos o impacto das variantes P.1 e P.2 usando simulações de dinâmica molecular para o complexo ACE2-RBD. Buscamos encontrar explicações teóricas das causas do aumento da transmissibilidade viral ou mesmo da gravidade dos sintomas, além do porquê haver convergência de determinadas mutações. Ao final, foram analisadas as flutuações atômicas mediante análises como RMSD e RMSF. Além disso, realizou-se decomposição da energia MM-PBSA envolvida na interação do complexo.

### Metodologia

Primeiramente, em decorrência da indisponibilidade de estruturas cristalográficas contendo as mutações da linhagem P.1 (K417T, E484K e N501Y) e P.2 (K417N, E484K, N501Y), foi necessário utilizar estruturas de referência a partir do banco de dados RCSB *Protein Data Bank* (<https://www.rcsb.org/>) [4] onde as substituições de aminoácidos foram aplicadas. A mutação *in silico* foi realizada no software PyMOL 2.3 [5] com o módulo "*Mutagenesis*" tendo o rotâmero com tensão estérica mais baixa definido automaticamente pelo software. A reconstrução de nova conformação de ACE2-RBD (PDB ID: 6M0J) [6] com mutações da P.1 e P.2 foi possível com a implementação do algoritmo de enumeração implícita Branch-and-Prune [7]. Nas instâncias considerou-se um conjunto arbitrário de distâncias inferiores a 6Å a partir da simulação de dados de RMN. Apenas a solução matematicamente válida e que também fosse validada pelo diagrama de Ramachandran foi submetida posteriormente às simulações de dinâmica molecular.

Realizou-se simulações atomísticas de dinâmica molecular no intervalo de 18 ns. Todas as proteínas foram previamente preparadas com o software em sua versão para fins acadêmicos Schrödinger Maestro 2020-4 utilizando o módulo "*Protein Preparation*". A geração de todos os arquivos de entrada foi feita utilizando o plugin QwikMD 1.3 [8] implementado na interface gráfica VMD 1.9.4.51a [9]. O complexo foi imerso em uma

caixa de solvatação com geometria cúbica e aplicação de PBC (condições periódicas de contorno) contendo moléculas de água descritas pelo modelo TIP3P [10], bem como a adição de íons  $\text{Na}^+$  e  $\text{Cl}^-$  para neutralizar o sistema em uma concentração fisiológica de  $0,15 \text{ mol} \cdot \text{L}^{-1}$ . Além disso, todos os arquivos de topologia foram gerados com o campo de força CHARMM36 [11]. O sistema foi minimizado usando 1000 passos mediante gradiente conjugado, onde depois houve aquecimento gradual de 60-300K sob o ensemble NPT e então o sistema foi equilibrado em NPT ao longo de 1 ns.

Por fim, a etapa de cálculo da trajetórias utilizou a integração de Verlet em um ciclo termodinâmico NPT com o tempo aproximado de 18 ns. O pistão estocástico da dinâmica de Langevin foi usado para controle de pressão sob 1 atm análogo a um barostato, enquanto um termostato de Nosé-Hoover manteve a temperatura em 300,0 K. Além disso, as interações eletrostáticas de longo alcance foram estimadas de acordo com o formalismo Particle-Mesh-Ewald (PME). Finalmente, o algoritmo de simulação NAMD3 (*Nanoscale Molecular Dynamics*) [12] foi usado para executar todas as simulações com aceleração da GPU Nvidia GTX 1050 de 2 GB com 640 núcleos CUDA em um passo de tempo de 2 fs.

Ao final, foram mensuradas as mudanças conformacionais na macromolécula em decorrência das variantes, através da evolução temporal RMSD e RMSF do deslocamento atômico em relação ao C $\alpha$  do complexo ACE2-RBD, onde o frame 0 foi adotado como referência. A análise de trajetórias foi possível com a biblioteca MDAAnalysis [13] implementada em Python. O cálculo do quantitativo total de ligações de hidrogênio foi feito usando um plugin presente no software VMD. Todos os valores SASA foram estimados para a média dos frames onde o algoritmo empregado também foi um plugin no VMD.

A decomposição dos valores de energia ( $\langle E_{\text{elec}} \rangle$ ,  $\langle E_{\text{vdw}} \rangle$  e  $\langle E_{\text{MM}} \rangle$ ) foram todos em relação à média dos frames de dinâmica molecular, onde utilizou-se o plugin NAMDenergy para os cálculos. A partir da área de solvatação média foi possível calcular o termo  $\langle E_{\text{solvatação apolar}} \rangle$  onde as constantes empíricas  $\gamma \approx 0,00542 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  e  $\beta \approx 0,92 \text{ kcal} \cdot \text{mol}^{-1}$  foram adotadas. A componente  $\langle G_{\text{PB}} \rangle = \langle E_{\text{solvatação polar}} \rangle$  de Poisson-Boltzmann foi estimado para o último quadro pelo método das diferenças finitas usando o algoritmo DelPhi v8.4.5 [14]. Como resultado da alta complexidade computacional e da incerteza inerente à estimativa do termo entrópico  $T \cdot \Delta S$ , não foi incluído na decomposição MM-PB(SA) desta pesquisa.

## Resultados e Discussão

Percebeu-se que 4 (quatro) análises (veja Tabela 1) corroboram a hipótese de maior estabilidade ACE2-RBD em consequência da P.1, dentre estas: Baixos valores médios de RMSD e RMSF, baixa exposição SASA e maior formação de ligações de Hidrogênio. Por outro lado, apenas 2 (duas) análises refletem a hipótese de maior instabilidade, sendo estas: Menor fração de contatos nativos e menor compactação medida pelo Raio de Giro. Em geral, embora as diferenças de flutuação possam ser parcialmente uma consequência de uma nova distribuição de velocidade associada à natureza estocástica de cada simulação realizada, as mutações da P.1 de um modo geral acarretaram de fato em maior estabilização do complexo.

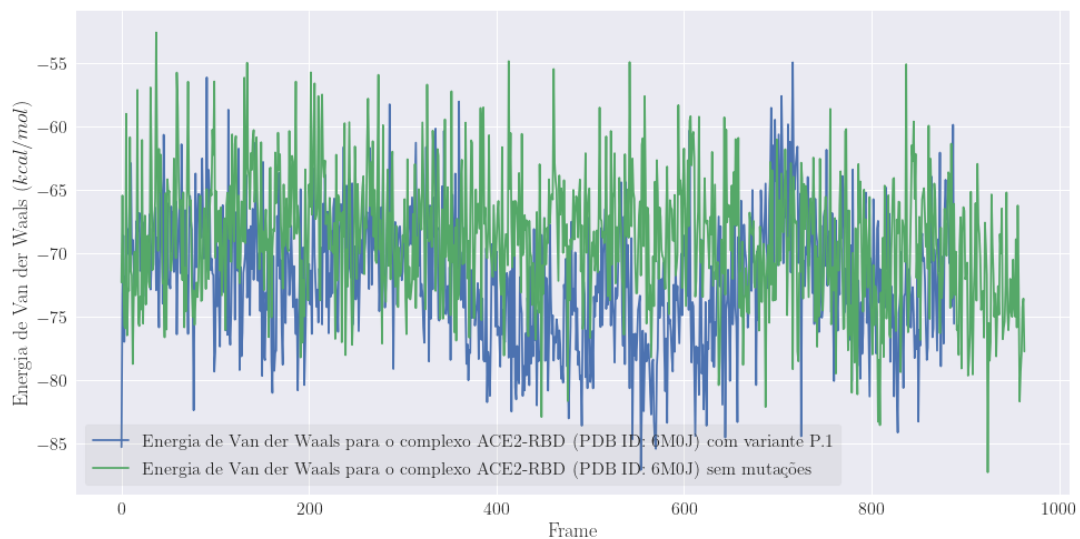
**Tabela 1:** Comparação entre os valores médios de alguns parâmetros resultantes da dinâmica molecular em 18 ns que caracterizam as alterações estruturais na proteína Spike na interação com o receptor ACE2 (PDB ID: 6MOJ).

Tipo de análise	Sem mutações	Linhagem P.1	Linhagem P.2
Média RMSD	(2,05 ± 0,32) Å	(1,76 ± 0,26) Å	(1,81 ± 0,34) Å
Média RMSF	(1,22 ± 0,51) Å	(1,06 ± 0,38) Å	(1,33 ± 0,76) Å
Ligações de Hidrogênio	197 ± 12	202 ± 12	203 ± 11
Média SASA	(378,55 ± 3,64) nm <sup>2</sup>	(375,04 ± 3,16) nm <sup>2</sup>	(373,35 ± 3,32) nm <sup>2</sup>
Fração de Contatos nativos	0,9918 ± 0,0015	0,9895 ± 0,0016	0,9877 ± 0,0019
Raio de Giro	(31,37 ± 0,30) Å	(31,64 ± 0,18) Å	(31,48 ± 0,18) Å

A maior estabilização no complexo ACE2-RBD como resultado de certas mutações pode ser uma explicação do porquê o vírus tem seguido uma evolução convergente como já reportado em alguns estudos experimentais [3], embora as causas até então são desconhecidas. Através das simulações deste trabalho, há indícios de existir a tendência a uma maior estabilidade estrutural nas mutações mais frequentes. Em outras palavras, as mutações que têm se repetido em várias cepas, como E484K e N501Y, tendem a ter maior estabilidade termodinâmica. A formação favorável e maior estabilidade do complexo ACE2-RBD em função da P.1 pode estar correlacionado a termos ainda menores da energia de Van der Waals ( $\langle E_{\text{vdw}} \rangle$ ) (veja Tabela 2 e Figura 1). Ao que tudo indica a contribuição de Van der Waals vem tendo um papel central para uma maior formação de ligações intermoleculares no complexo contendo a P.1.

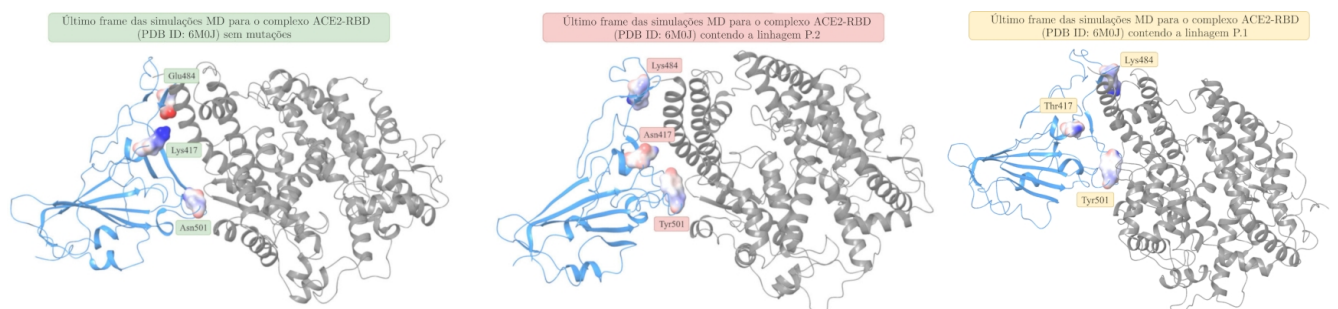
**Tabela 2:** Todos os valores da decomposição de energia MM-PB(SA) referem-se ao complexo ACE2-RBD (PDB ID: 6M0J) em simulações de 18 ns. Os valores de energia estão na unidade kcal · mol<sup>-1</sup>. O termo para as interações covalentes ( $E_{int}$ ) foi considerado aproximadamente zero nas análises.

Termo	Sem mutações	Linhagem P.1	Linhagem P.2
$E_{elec}$	-185,57 ± 40,42	-66,51 ± 31,17	-45,91 ± 25,62
$E_{vdw}$	-68,85 ± 5,05	-72,07 ± 5,08	-74,15 ± 4,91
$E_{MM}$	-254,42 ± 40,90	-138,59 ± 31,26	-120,06 ± 26,43
$G_{solvatação\ polar}$	-2628,61	-2663,48	-2570,16
$G_{solvatação\ apolar}$	206,10	204,2	203,28
$G_{solvatação\ total}$	-2422,51	-2459,29	-2366,88
Entalpia	-2676,93	-2597,88	-2486,94



**Figura 1:** Gráfico comparativo do impacto da linhagem P.1 na energia de Van der Waals ( $E_{vdw}$ ) na interação do complexo ACE2-RBD (PDB ID: 6M0J).

No aspecto qualitativo (veja Figura 2) percebe-se que a linhagem P.1 tornou a folha- $\beta$  formada entre as mutações K417N e N501Y com um menor comprimento comparado à estrutura ausente de mutações, além de ter havido o desaparecimento de uma  $\alpha$ -hélice, fatores estruturais que podem estar correlacionados com a maior transmissibilidade da variante P.1.



**Figura 2:** As superfícies de distribuição eletrostática correspondem aos aminoácidos que foram afetados pela P.1/P.2, de modo que foram geradas com o algoritmo APBS a partir dos últimos frames das respectivas simulações de dinâmica molecular para o complexo ACE2-RBD (PDB ID: 6M0J). A visualização foi por intermédio do software Schrödinger Maestro 2020-4.

Em consequência da interação ACE2-RBD ser um processo dependente da temperatura, conforme descoberto na pesquisa [15], então as premissas  $\Delta H < 0$  e  $\Delta S < 0$  são válidas. Além disso, sabe-se que os coronavírus tendem a se replicar com maior probabilidade diante da sazonalidade no inverno, o que indica valores ainda mais negativos de  $\Delta G$ . Consequentemente, pode-se concluir que a linhagem P.1, sendo mais estável, terá menos dificuldade em superar a penalidade entrópica e portanto mais resistente ao aumento de temperatura e maior afinidade de interação.

Por fim é importante lembrar que os impactos das linhagens P.1 e P.2 se tornarão mais nítidos à medida que houver a repetição das simulações ou aumento no intervalo de tempo, o que geraria resultados mais conclusivos, convergência das flutuações e maior reprodutibilidade. Por isso dentre as perspectivas futuras é justamente o prosseguimento às simulações para aprofundar o entendimento à nível teórico dos impactos das variantes emergentes do SARS-CoV-2.

### Conclusões

Em geral, as linhagens P.1 e P.2 proporcionaram maior estabilização do complexo ACE2-RBD corroborado por 4 (quatro) análises da dinâmica molecular, os valores médios RMSD e RMSF, maior formação de ligações de Hidrogênio e menor exposição ao solvente. Consequentemente, a priori, poderão não desaparecer devido à pressão evolutiva, visto que beneficiam o vírus, e talvez por isso haja uma convergência de certas mutações como E484K e N501Y. Por fim, espera-se com esses resultados auxiliar na compreensão teórica dos impactos das variantes emergentes no Brasil. Por esse motivo, é essencial atualizar as vacinas o mais rápido possível para minimizar os efeitos das mutações não sinônimas atuais e futuras especificamente na glicoproteína Spike.

### Referências

- [1] F. Naveca *et al.*, “COVID-19 epidemic in the Brazilian state of Amazonas was driven by long-term persistence of endemic SARS-CoV-2 lineages and the recent emergence of the new Variant of Concern P.1,” 2021, doi: <https://doi.org/10.21203/rs.3.rs-275494/v1>.
- [2] N. R. Faria *et al.*, “Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil,” *Science* (80- ), 2021, doi: 10.1126/science.abh2644.
- [3] S. A. Kemp *et al.*, “Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70,” *bioRxiv*, 2021.
- [4] H. M. Berman *et al.*, “The Protein Data Bank,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/nar/28.1.235.
- [5] L. L. C. Schrödinger and W. DeLano, “PyMOL.” 2020, [Online]. Available: <http://www.pymol.org/pymol>.
- [6] J. Lan *et al.*, “Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor,” *Nature*, vol. 581, no. 7807, pp. 215–220, 2020, doi: 10.1038/s41586-020-2180-5.
- [7] G. Lin, D. Xu, Z.-Z. Chen, T. Jiang, J. Wen, and Y. Xu, “An Efficient Branch-and-Bound Algorithm for the Assignment of Protein Backbone NMR Peaks,” in *Proceedings of the IEEE Computer Society Conference on Bioinformatics*, 2002, p. 165.
- [8] J. V Ribeiro *et al.*, “QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts,” *Sci. Rep.*, vol. 6, no. 1, p. 26536, 2016, doi: 10.1038/srep26536.
- [9] W. Humphrey, A. Dalke, and K. Schulten, “VMD: Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996, doi: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [10] P. Mark and L. Nilsson, “Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K,” *J. Phys. Chem. A*, vol. 105, no. 43, pp. 9954–9960, Nov. 2001, doi: 10.1021/jp003020w.
- [11] J. Huang and A. D. MacKerell Jr, “CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data,” *J. Comput. Chem.*, vol. 34, no. 25, pp. 2135–2145, Sep. 2013, doi: <https://doi.org/10.1002/jcc.23354>.
- [12] J. C. Phillips *et al.*, “Scalable molecular dynamics on CPU and GPU architectures with NAMD,” *J. Chem. Phys.*, vol. 153, no. 4, p. 44130, 2020, doi: 10.1063/5.0014475.
- [13] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations,” *J. Comput. Chem.*, vol. 32, no. 10, pp. 2319–2327, Jul. 2011, doi: <https://doi.org/10.1002/jcc.21787>.
- [14] L. Li *et al.*, “DelPhi: a comprehensive suite for DelPhi software and associated resources,” *BMC Biophys.*, vol. 5, no. 1, p. 9, 2012, doi: 10.1186/2046-1682-5-9.
- [15] J. He, H. Tao, Y. Yan, S. Y. Huang, and Y. Xiao, “Molecular mechanism of evolution and human infection with SARS-CoV-2,” *Viruses*, vol. 12, no. 4, Apr. 2020, doi: 10.3390/v12040428.