



# Investigação de uma proteína de SARS-CoV-2 por técnicas computacionais

**Palavras-chave:** evolução enzimática, dinâmica conformacional, dinâmica molecular.

**Luiz De Martino Costa** (UNICAMP – Universidade Estadual de Campinas | [luizdmcosta@gmail.com](mailto:luizdmcosta@gmail.com))

**Prof. Dr. Ronaldo Junio de Oliveira** (UFTM – Universidade Federal do Triângulo Mineiro)

**Prof.<sup>a</sup> Dra. Denize Cristina Favaro** (UNICAMP – Universidade Estadual de Campinas)

---

## INTRODUÇÃO:

Os últimos dois anos estão sendo marcados pela pandemia do vírus SARS-CoV-2 e, atualmente, suas variantes, já acarretando até o dia 22/08/2021 um total de 210 milhões de casos confirmados e 4,4 milhões de mortes no mundo inteiro, tendo o Brasil 9,8 % dos casos e 13,0 % das mortes (OMS, 2021). Além disso, como há a possibilidade da Covid-19 se tornar uma doença sazonal se não houver o devido controle (ONU, 2021) e como é o terceiro coronavírus com implicações severas no sistema respiratório (SANTOS, 2020), a compreensão detalhada sobre seu funcionamento é extremamente necessária.

Uma das proteínas com atuação menos conhecida do SARS-CoV-2 é a nsp 9 (*nsp: non-structural proteins*), uma proteína homodimérica que apresenta 97,3 % de homologia de sequência com a nsp9 de SARS-CoV-1 e possui um domínio  $\alpha$ -hélice GxxxG considerado essencial para a replicação do vírus (MIKNIS et al., 2009). Ensaios *in vitro* mostram interações da nsp 9 com RNA e DNA de fita simples (EGLOFF et al., 2004), além disso, foi relatado que a nsp9 participa de forma destrutiva no processo de enovelamento de proteínas nos ribossomos (LITTER et al., 2020). Adicionalmente, cogita-se que a nsp 9 participa do processo de transferência de nucleosídeo de monofosfato vinda da nsp12 usando os resíduos N1 e N2 (YAN et al., 2021; SLANINA et al., 2021).

## OBJETIVOS:

Caracterizar interações e resíduos importantes da proteína nsp9 de SARS-CoV-2 utilizando técnicas computacionais. Para isso foram realizados:

- Dinâmica molecular da nsp9 de SARS-CoV-2 em solvente explícito;
- Simulações de “docking” molecular da nsp9 com a nsp12 e posteriormente da proteína de interesse com RNA segundo dados de Banerjee et al. (2020) e Yan et al. (2021);
  - “Redocking” com peptídeo citado na literatura por Litter et al. (2020) para definir se a técnica traz resultados coerentes com dados experimentais;
- Comparação por  $\Delta G$  de solvatação da nsp9 de SARS-CoV-2 com a de SARS-CoV-1, bem como das mutantes G100A, G104A, G100A G104A, N1A, N2A e N1A N2A com o intuito de investigar resíduos importantes, funcionalmente e estruturalmente, na estabilidade da estrutura;
- Dinâmica molecular baseada em modelos estruturais da nsp9 de SARS-CoV-2, de SARS-CoV-1 e mutantes a fim de se investigar distinções energéticas no sentido tempo e temperatura baseada na mudança do número de contatos realizados.

## METODOLOGIA:

Preparação das proteínas: as proteínas foram produzidas a partir da sequência da nsp9 de SARS-Cov-2 pelo preditor de estrutura *online I-Tasser* (YANG; ZHANG, 2015) visto que as estruturas cristalográficas dela e de SARS-CoV-1 não constam com os resíduos N1 e N2 bem representados, apesar de serem importantes em interações

relatadas da proteína. Estruturas das mutantes de SARS-CoV-2 e SARS-CoV-1 foram geradas usando uma ferramenta de mutação do *software Chimera* (PETTERSEN et al., 2004). Por fim se gerou as estruturas de trabalho fazendo minimização energética pelo servidor *online Yassara* (KRIEGER et al., 2009).

Estudos de interação do tipo "docking" molecular: utilizou-se o servidor *online HADDOCK 2.4* (ZUNDERT et al., 2016) para a simulação de interações.

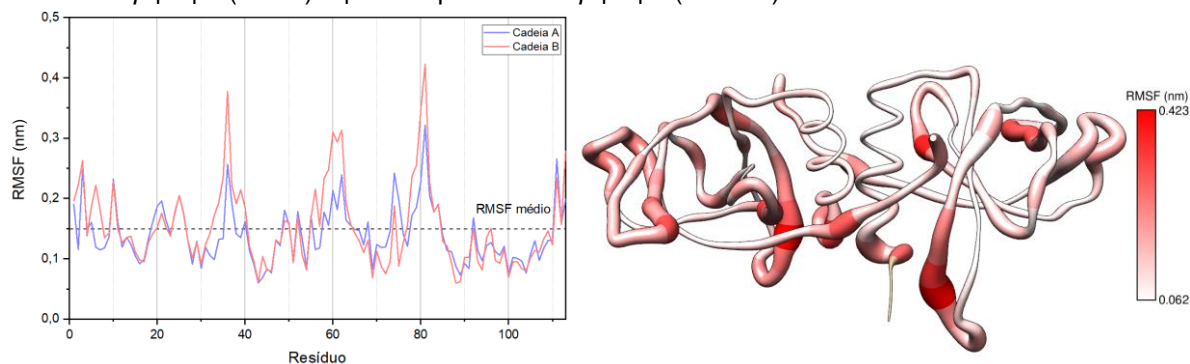
Comparação energética de estruturas: utilizou-se o servidor *online PDBE PISA* (KRISINEL; HENRICK, 2007) para análise baseada em energia de solvatação das estruturas de interesse.

#### Dinâmica Molecular:

- Solvente Explícito: fez-se a simulação de flutuação da nsp9-SARS-Cov-2 em solvente explícito com o *software Gromacs* (SPOEL et al., 2005) com o campo de força GROMOS54a7 (CHRISTEN et al., 2005) por simulações com passo de 2 fs para um tempo total de 200 ns. O sistema foi simulado na temperatura de 300 K, na pressão de 1,0 bar em uma caixa dodecaédrica rômica com a distância de 1 nm da extremidade da proteína. O sistema foi preenchido com água modelo spc216 e concentração de 150 mM de NaCl. Antes da simulação de flutuação, o sistema foi minimizado em vácuo e em solvente com e sem restrição dos movimentos da proteína e posteriormente equilibrado por NPT e NVT.
- Modelo Baseado em Estrutura: usou-se o pacote SMOG 2.3 (NOEL et al., 2016) para a criação do modelo de estruturas e se utilizou o *software Gromacs* (SPOEL et al., 2015) para as simulações com passo de 2 fs em caixas cúbicas. Foram feitas duas bateladas de experimentos: uma primeira variando de 80 K a 150 K por 2 ns com distância de 2 nm da extremidade da proteína, enquanto a outra varia de 100 K a 130 K por 6 ns com distância da extremidade de 4 nm. Em ambos os casos, as variações de temperatura foram feitas com 8 intervalos equivalentes. Os números de contatos pelo tempo foram analisados com *script contacts-python* (FREITAS, 2021).

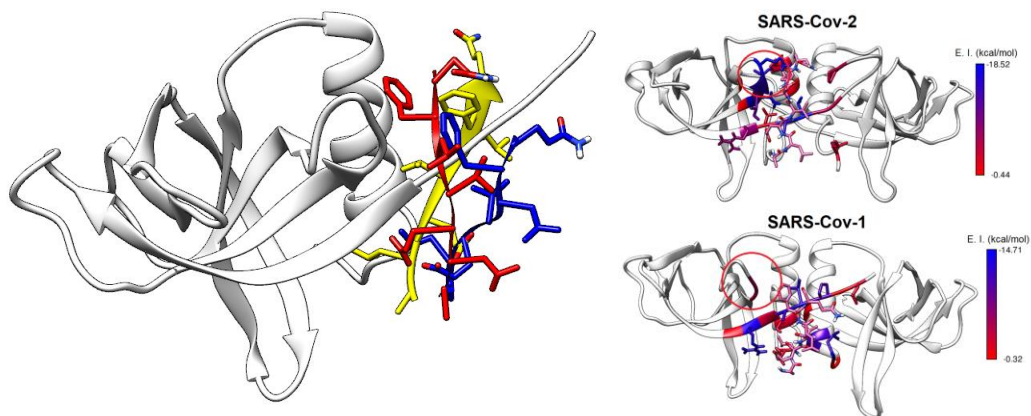
## RESULTADOS E DISCUSSÃO:

Dinâmica Molecular em Solvente Explícito: A simulação de flutuação da nsp9 de SARS-CoV-2 apresentou um comportamento estável sem mudanças conformacionais durante o tempo de simulação, o que foi visto pela estabilidade do valor do raio de giro. Essa observação posteriormente, foi verificada por *Clusterização* e por *Superfície de Energia Livre* que forneceram o mesmo tipo de conformação. Fazendo uma análise da flutuação quadrática média (RMSF) dos resíduos pelos nos últimos 50 ns, presente na **Figura 1**, percebe-se uma relativa rigidez no domínio GxxxG, enquanto é observada uma maior flexibilidade nas regiões: N-terminal (resíduos 1 – 10),  $\beta 2$  com parte do *loop*  $\beta 2/\beta 3$  (34 – 39),  $\beta 4$  com parte do *loop*  $\beta 4/\beta 5$  (58-63) e  $\beta 6$  com parte do *loop*  $\beta 6/\beta 7$  (78 – 83).



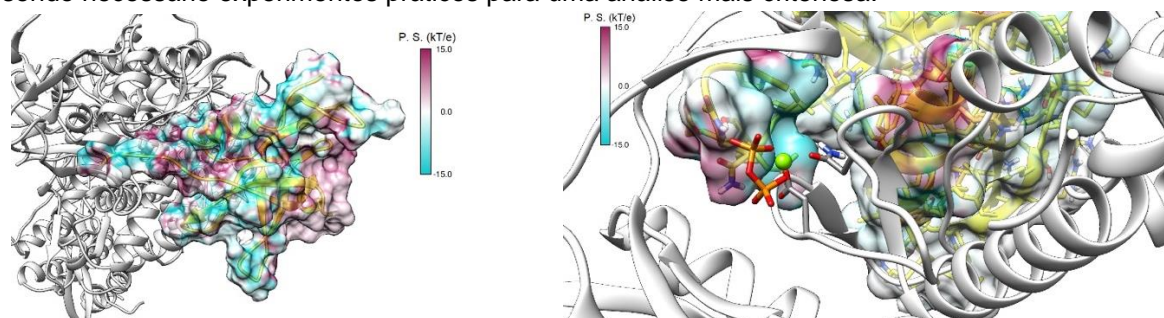
**Figura 1:** resultados de flexibilidade da nsp9 nos últimos 50 ns. À esquerda: gráfico dos valores de RMSF por resíduo e cadeia. À direita: representação dos dados de RMSF.

Teste de "Redocking": usando como experimento inicial de aferição da técnica de "docking" molecular para investigar interações da nsp9 de SARS-CoV-2, escolheu-se a proteína com peptídeo no trabalho de Litter et al. (2020). Montando a estrutura apenas do peptídeo com sequência ALEVLQ (PDB: 6W9Q), pode-se ver que o algoritmo do HADDOCK 2.4 conseguiu prever tanto para SARS-CoV-2 quanto para SARS-CoV-1 as zonas de interação descritas: as com o domínio GxxxG e as com a região N-terminal. Além disso, ambas poses ocupam espaços próximos do peptídeo e apresentam significativas energias de interação, como pode ser visto pela **Figura 2**. As maiores diferenças das poses provavelmente se devem ao fato de a estrutura original do peptídeo ter sido feita com ele incluso em sequência da região N-terminal, afetando ligeiramente a geometria da região. Dessa forma, a técnica de "docking" se viu aceitável para moléculas com interação ainda relativamente desconhecidas.

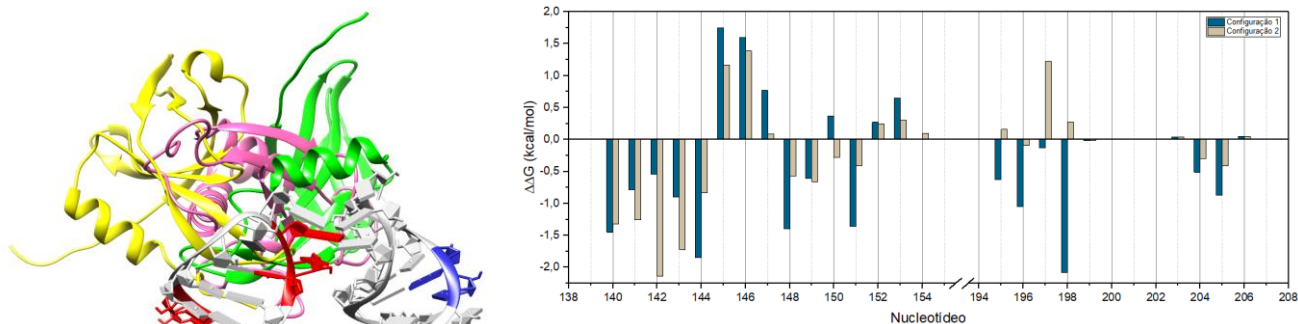


**Figura 2:** resultados do processo de “redocking”. À esquerda: sobreposição das melhores poses (peptídeo, SARS-CoV-2 e SARS-CoV-1). À direita: representação dos dados de energia de interação (E. I.) dos “redockings”.

**Análises de interação via “docking” da nsp9 de SARS-CoV-2 com outros sistemas (nsp12 e RNA):** a interação da nsp9 monomérica com a nsp12 foi feita baseada no artigo de Yan et al. (2021). Usando o servidor online APBS (DOLINSKY et al., 2004), foi possível observar zonas de potencial eletrostático da região N-terminal e  $\alpha$ -hélice da nsp9 em uma cavidade contendo pirofosfato de magnésio da nsp12, como pode ser visto na **Figura 3**. Na cavidade, a extremidade do N-terminal (resíduos N1 e N2) está bem próxima ( $\sim 4$  Å) desse composto, o que é coerente com os trabalhos de Slanina et al. (2021) e Yan et al. (2021) nos quais discutem sobre a influência da nsp9 no processo de reprodução do vírus. Segundo o primeiro autor, os dois resíduos de asparagina são importantes para o processo de transferência de monofosfato nucleosídico em uma das etapas do processo. Para as interações com RNA, utilizou-se como base as regiões de nucleotídeos marcadas no trabalho de Banerjee et al. (2020), como pode ser visto em vermelho na **Figura 4** (nucleotídeos 142-143 e 149-151). Como o trabalho não apresentava a estrutura com a nsp9 e não há semelhanças estruturais e sequenciais entre a nsp9 e a SRP19, proteína que a nsp9 ocuparia o local, foi necessário obter um preditor de interfaces para uma estimativa inicial dos resíduos. Para isso se utilizou o servidor online PRIME-3D2D (XIE et al., 2020), sendo então possível o “docking”. Os resultados apontaram duas potenciais configurações, Configuração 1 e Configuração 2, mostradas na **Figura 4**. Fazendo comparações estruturais das energias dos nucleotídeos em comparação à proteína SRP19, ambas se mostraram adequadas para o sistema, contribuindo mais para a estabilização do RNA, sendo necessário experimentos práticos para uma análise mais criteriosa.



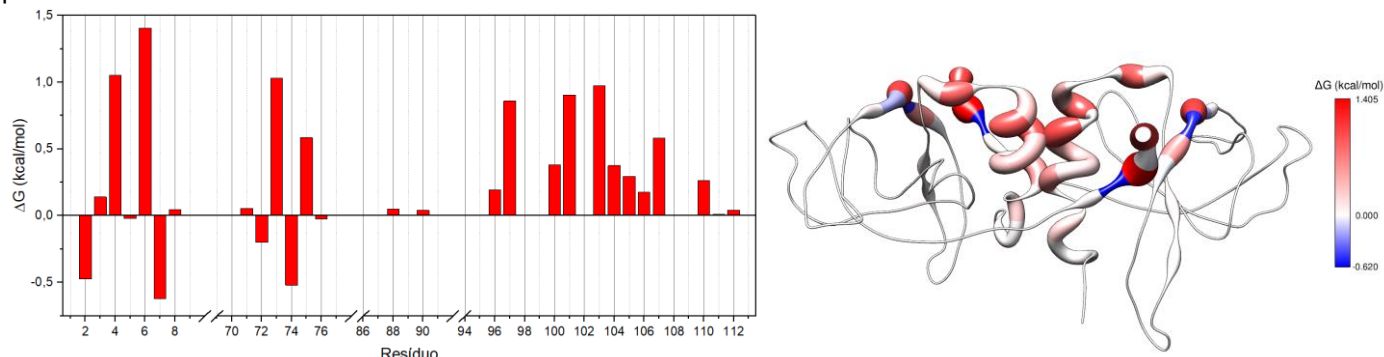
**Figura 3:** estruturas com superfície baseada em potencial de superfície (P.S.) da nsp9 de SARS-CoV-2 com nsp12 de forma geral (à esquerda) e com foco na cavidade ocupada pela nsp9 (à direita).



**Figura 4:** resultados do “docking” de RNA. À esquerda: estrutura da SRP19 (PDB: 1MFQ) superimposta com melhores poses da nsp9 de SARS-CoV-2 usando como referência RNA (Configuração 1, Configuração 2, SRP19, RNA previsto de interagir com nsp9 e nsp8). À direita: gráfico das diferenças das contribuições energéticas entre as formações propostas contra SRP19.

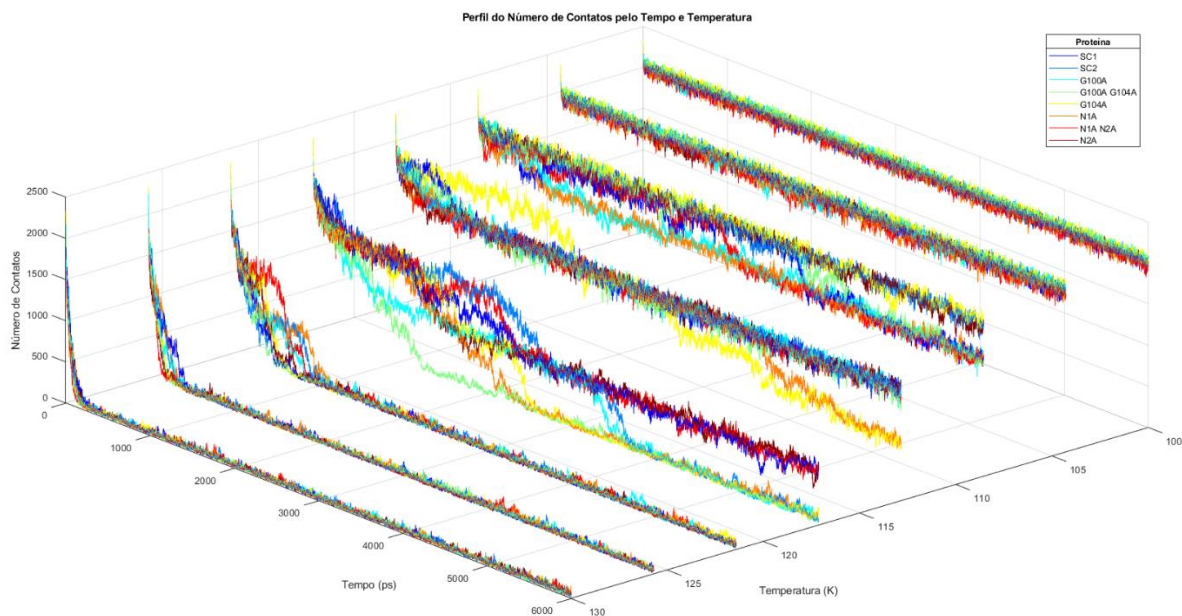


Comparação energética entre estruturas por energia de solvatação: após realizadas as análises de estrutura pelo servidor PDBe PISA, uma primeira observação é a formação de 5 a 9 ligações de hidrogênio entre os resíduos 5-8 (região N-terminal) de uma cadeia com o resíduo 77 (região  $\beta 6$ ) da outra, bem como a formação de 4 pontes salinas entre os resíduos 6 e 77 para todas as estruturas de nsp9. Além disso, foi visto que os resíduos 2-7, 71-76, 88, 90, 96-97, 100-101, 103-107, 110 e 112 contribuem para energia de solvatação nsp9 de SARS-CoV-2, conforme pode ser observado pela **Figura 5**. Dessa forma, o domínio  $\alpha$  possui também uma observável influência na estabilidade da proteína.



**Figura 5:** resultados das contribuições energéticas de cada resíduo da nsp9 de SARS-CoV-2. À esquerda: gráfico dos valores de  $\Delta G$  por resíduo. À direita: representação dos dados de  $\Delta G$ .

Dinâmica molecular baseada em estrutura: as proteínas de forma geral apresentaram distinção nos seus comportamentos pelas variáveis número de contatos, tempo e temperatura como pode ser visto na **Figura 6**. Especificamente as mutantes G100A, G104A, G100A G104A (todas as mutações na região  $\alpha$ -hélice) e N1A mostraram uma queda do número de contatos mais acentuada pelo tempo para as mesmas condições de temperatura, indicando que necessitam de uma menor energia para desenovelarem. Nos outros casos, não pode ser observada uma diferença notável do perfil em comparação à nativa ou à nsp9 de SARS-CoV-1. Além disso, ainda não se pode ser afirmado nenhuma informação sobre barreiras energéticas entre as energias de estados diméricos, monoméricos e desenovelados, apesar dos perfis obtidos aparentarem a formação de níveis estáveis com degraus.



**Figura 6:** gráfico comparativo entre proteínas pelo tempo, temperatura e número de contatos.

## CONCLUSÃO:

A nsp9 do vírus SARS-CoV-2 e homólogos ainda é uma pesquisa em desenvolvimento, mas importante para o entendimento do funcionamento desse tipo de vírus. Essa investigação, baseada em técnicas computacionais, permitiu-nos criar *insights* sobre essa proteína e como abordar esse tipo de problema, mas ainda são observações que

devem ser expandidas e validadas na prática. Foi visto que a região GxxxG se mostrou rígida pela técnica de dinâmica molecular, mas importante no ponto de vista energético e de estabilidade com base na análise das energias de solvatação e da dinâmica baseada em estrutura. Além disso, também se observou que a região N-terminal possui bastante contribuição energética na interação com a nsp12, tendo nesse caso destaque os resíduos N1 e N2. Além disso, foram feitas duas propostas de configurações para regiões de interação com o RNA para o início de uma investigação mais completa.

---

## BIBLIOGRAFIA

- BANERJEE, A. K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. **Cell**, EUA, v. 183, p. 1-15, 25 nov. 2020.
- CHRISTEN, M. *et al.* The GROMOS software for biomolecular simulation: GROMOS05. **J Comput Chem**, EUA, 26, 1719-1751, 2005.
- DOLINSKY T.J. *et al.* PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. **Nucleic Acids Research**, 32, W665-W667, 2004.
- EGLOFF, M. *et al.* The severe acute respiratory syndrome-coronavirus replicative protein Nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. **PNAS**, EUA, v. 101, p. 3792-3796, 2004.
- FREITAS, Frederico Campos. **Contacts-python**. Brasil, 2021. Disponível em: <https://github.com/fredcfreitas/contacts-python>. Acesso em: 16 jul. 2021.
- KRIEGER, Elmar *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. **Proteins**, EUA, v. 77, p. 114-122, 2009.
- KRISSINEL, Evgeny; HENRICK, Kim. Inference of macromolecular assemblies from crystalline state. **J. Mol. Biol.**, Holanda, v. 372, 2007
- LITTLER, D. R. *et al.* Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9. **IScience**, EUA, v. 23, 24 jul. 2020.
- MIKNIS, Z. J. *et al.* Severe Acute Respiratory Syndrome Coronavirus Nsp9 Dimerization Is Essential for Efficient Viral Growth. **Journal of Virology**, EUA, v. 83, p. 3007–3018, 2009.
- NAÇÕES UNIDAS. Relatório examina relação de fatores meteorológicos, qualidade do ar e Covid-19. In: **ONU News**. [S. l.], 18 mar. 2021. Disponível em: <https://news.un.org/pt/story/2021/03/1744912> . Acesso em: 20 ago. 2021.
- NOEL, Jeffrey K. *et al.* SMOG 2: A Versatile Software Package for Generating Structure-Based Models. **PLOS Computational Biology**, EUA, v. 12, 2016.
- ORGANIZAÇÃO MUNDIAL DA SAÚDE. **WHO Coronavirus (COVID-19) Dashboard**. [S. l.], Desde 2020 2020. Disponível em: <https://covid19.who.int/> . Acesso em: 22 ago. 2021.
- PETTERSEN, Eric F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. **J. Comput. Chem.**, EUA, v. 25, 2004.
- SANTOS, Maria Tereza. **As diferenças e semelhanças entre outros coronavírus e o Sars-CoV-2**. Brasil, 27 jul. 2020. Disponível em: <https://saude.abril.com.br/medicina/as-diferencas-e-semelhancas-entre-o-sars-cov-2-e-outros-coronavirus/>. Acesso em: 20 ago. 2021.
- SLANINA, H. *et al.* Coronavirus replication–transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. **PNAS**, EUA, v. 118, 2021.
- SPOEL, D. *et al.* GROMACS: fast, flexible, and free. **J Comput Chem** 26, 1701-1718, 2005.
- XIE, Juan *et al.* PRIME-3D2D is a 3D2D model to predict binding sites of protein–RNA interaction. **Communications Biology**, EUA, 2020.
- YAN, L. *et al.* Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. **Cell**, EUA, v. 184, p. 184-193, 7 jan. 2021.
- YANG J., ZHANG Y. I-TASSER server: new development for protein structure and function predictions. **Nucleic Acids Research**, 43, W174-W181, 2015.
- ZHANG, C. *et al.* Structural basis for the multimerization of nonstructural protein Nsp9 from SARS-CoV-2. **Molecular Biomedicine**, Suíça, 2020.
- ZUNDERT, G. C. P. *et al.* The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. **J. Mol. Biol.**, EUA, 428, 720-725, 2015.