

PREDIÇÃO E ANÁLISE FÍSICO-QUÍMICA *IN SILICO* DE COMPOSTOS BIOATIVOS CONTRA A PROTEÍNA SPIKE DA VARIANTE ÔMICRON – SARS-COV-2

Elcio Ferreira Santana¹
Bruna Chaves Amorim²
Wander José dos Reis³
Adriano Carlos Soares⁴

ÁREA DE CONHECIMENTO: Ciências da Saúde

RESUMO

A ômicron ou B.1.1.529 é uma variante do SARS-CoV-2 que é altamente contagiosa, aumentando assim o número de casos no mundo. Em sua membrana há a proteína *spike* (espigão) que serve como ligante para se conectar às células, causando assim maior probabilidade de ocorrerem doenças pulmonares, o que aumenta o risco de vida da população. Neste estudo, se fez a análise e predição físico-química de compostos bioativos a partir de *softwares* de bioinformática. Foram analisadas 22 proteínas que apresentaram domínios funcionais com identidade favorável ao ligante, obtido no *Protein Data Bank* (Banco de Dados de Proteínas), referente a *spike*. Após as análises do índice de instabilidade e *Grand average of hydropathicity* (Grande média de hidropatia) (*GRAVY*), no *software* suíço *ProtParam* (Parâmetros Proteicos) foram refinados 8 compostos candidatos a vacina. Logo, foi demonstrado *in silico* que os parâmetros físico-químicos e as análises confirmadas demonstram grande potencial bioativo contra a variante ômicron, podendo ser realizados estudos no sentido de desenvolver um medicamento ou vacina contra o vírus em questão.

PALAVRA CHAVE: variante ômicron; análise físico-química; bioinformática.

¹ Acadêmico em Farmácia no Centro Universitário Univértix - Bolsista do PIBIC/UNIVÉRTIX.

² Farmacêutica Generalista, Mestre em Ciências Naturais e da Saúde, Especialista em Docência do Ensino Superior, Professora e Coordenadora do Curso de Bacharelado em Farmácia da Faculdade Vértice – UNIVÉRTIX - Matipó.

³ Licenciado em Ciências Biológicas (UNEC), Especialista em Avaliação de Risco e Perícia Ambiental (UNEC), Pós-graduando em Docência do Ensino Superior (UNIFAVENI), Professor de Biologia no Centro Educacional de Matipó, Professor de Biologia no Colégio Losango de Raul Soares, Professor no curso de Biologia da UNIFAVENI.

⁴ Farmacêutico-Bioquímico (UFOP); Cirurgião Dentista (UNIVÉRTIX); Doutor em Bioquímica Aplicada (Biotecnologia) (UFV); Mestre em Ciências Naturais e da Saúde (UNEC); Especialista em Docência do Ensino Superior (UCAM, RJ), Especialista em Disfunção Temporomandibular e Dor Orofacial (UniBF, Paraná). Professor dos cursos de Farmácia, Psicologia, Enfermagem, Medicina e Odontologia do Centro Universitário Vértice – UNIVÉRTIX.

INTRODUÇÃO

No ano de 2019 foi detectado uma nova linhagem do coronavírus que se hospedava em morcegos, na cidade de *Wuhan* localizada na China, esse vírus pela sua alta taxa de transmissão causou uma pandemia que acarretou principalmente doenças pulmonares, renais e cardiovasculares (KHAN *et al.*, 2020).

Dessa forma, algumas variantes surgiram e a ômicron foi a mais preocupante devido a sua alta taxa de infecção e pelo fato de não haverem vacinas ou medicamentos específicos para combatê-la. Esse fator levou a grandes ondas de casos confirmados no mundo. A virulência dessa variante é devida a presença de 15 mutações no domínio de ligação ao receptor na proteína *spike* o que leva a uma maior ligação a Enzima Conversora de Angiotensina 2 (ECA2), localizada no pulmão. Essa ligação, ainda não havia sido descrita na literatura em outras linhagens de SARS-CoV. (ALKHATIB *et al.*, 2022).

No entanto, mesmo com vários avanços e colaborações na produção de vacinas no mundo, ainda não há um princípio ativo definitivo onde venha combater diretamente o vírus, visto que, o desenvolvimento dos tratamentos farmacológicos requer muito tempo e dinheiro (KHAN *et al.*, 2020). O uso de *softwares* de bioinformática para predição *in silico* de compostos bioativos é uma abordagem sofisticada, muito bem detalhada e que demanda vários estudos que validam a partir de cálculos computacionais dos parâmetros envolvidos confrontados com os dados experimentais validados e publicados na literatura. Esses dão suporte para a construção de *softwares* economizando recursos financeiros e tempo na pesquisa científica de novos compostos bioativos (KAUR, MADGULKAR, BHALEKAR, ASGAONKAR, 2019).

Para o estudo *in silico*, os parâmetros são obtidos a partir de plataformas científicas on line, por exemplo, o software suíço *ProtParam (Protein Parametres)* (Parâmetros Proteicos), o *Expasy (Expert Protein Analysis System)* (Sistema Especialista de Análise de Proteínas) (<https://web.expasy.org/protparam/>). Esses, foram desenvolvidos para obtenção de dados físico-químicos por meio de sequências de aminoácidos e/ou sequências depositadas no banco de dados mantido pelo Instituto Suíço de Bioinformática (GASTEIGER *et al.*, 2005).

O estudo da instabilidade de proteínas a partir de dipeptídeos mostra que o peso relacionado a sua cadeia de resíduos de aminoácidos pode demonstrar sua estabilidade ou instabilidade, isso leva a uma predição de sua estabilidade *in vivo* a partir de sua sequência de aminoácidos (GURUPRASAD, REDDY, PANDIT, 1990). Um dado muito importante é a medida de hidropatia de uma proteína, calculando a solubilidade de proteínas, o que ajuda na circulação do mesmo no organismo; o uso do cálculo de *Grand average of hydropathicity (Grande média de hidropatia) GRAVY* para fazer essa predição foi relacionado com dados obtidos em cristalografia (KYTE, DOOLITTLE, 1982).

Logo, o objetivo deste estudo é a realização da predição e análise físico-química *in silico* de compostos bioativos contra a proteína *spike* da variante ômicron do SARS CoV-2 a partir de *softwares* de bioinformática.

FUNDAMENTAÇÃO TEÓRICA

SARS-COV-2 E A PROTEÍNA SPIKE

A doença causada pelo novo coronavírus (COVID-19) foi denominada síndrome respiratória aguda grave. Essa é da família *Betacoronavírus*, que possui fácil infecção e transmissibilidade. A fisiopatologia conhecida até o momento é que, o vírus para entrar na célula infectada possui em seu capsídeo a proteína *spike* (espigão). Esta se liga a Enzima Conversora de Angiotensina 2 (ACE2), presente nos seres humanos (BARRETO-VIEIRA *et al.*, 2022). A presença da proteína *spike* leva a um aumento na taxa de infecção. Essa proteína resulta de uma mutação de proteínas do capsídeo viral levando a um potencial resistência a anticorpos do hospedeiro. Fato que transformou essa mutação em uma preocupação mundial, pois essas aparentemente, levam a maior afinidade ao receptor humano ACE2 e consequente internalização do vírus nas células (HE *et al.*, 2021). E ainda, não há medicamentos ou vacinas que ataquem diretamente o vírus, sendo necessária a busca por moléculas bioativas que venham a apresentar uma capacidade terapêutica contra essa nova linhagem de vírus (BARRETO-VIEIRA *et al.*, 2022).

SOFTWARES DE BIOINFORMÁTICA

O uso de *softwares* de bioinformática é muito importante devido a sua praticidade, pois os mesmos estão hospedados em bancos de dados de alto processamento e segurança, o que leva a uma economia de recursos, pois se baseiam em métodos computacionais validados cientificamente levando a uma economia de tempo. Sendo assim, é possível analisar diversas moléculas bioativas candidatas a novos fármacos poupando de pesquisa e recursos, e principalmente tempo (SOARES, 2015). Alguns bancos de dados com informações depositadas: *Protein Data Bank* (Banco de Dados de Proteínas) (<https://www.rcsb.org/>), *NCBI* (National Center for Biotechnology Information) (Centro Nacional de Informações sobre Biotecnologia) (<https://www.ncbi.nlm.nih.gov/>), *UniprotKB* (<https://www.uniprot.org/help/uniprotkb>) e *BLAST* (*Basic Local Alignment Search Tool*) (Ferramenta básica de pesquisa de alinhamento local) (<https://www.uniprot.org/blast/>) possuem as sequências de aminoácidos necessárias (TANTIVITAYAKUL *et al.*, 2020). A análise filogenética é importante para a visão da diferença entre as diferentes linhagens de compostos bioativos, para isso o *software MEGA11* (*Molecular Evolutionary Genetics Analysis Version 11*) (Análise de Genética Evolutiva Molecular Versão 11) (<https://www.megasoftware.net/>) é capaz de fazer esses cálculos precisos fornecendo informações importantes sobre desenvolvimento e estudo de biomoléculas (SOARES, 2015).

O software suíço *ProtParam* (*Protein Parameters*) (Parâmetros Proteicos) *ExPasy* (*Expert Protein Analysis System*) (Sistema Especialista de Análise de Proteínas) (<https://web.expasy.org/protparam/>) é utilizado para cálculos de parâmetros físico-químicos de índice de instabilidade e *GRAVY* (DUTTA, BANERJEE, CHAKRABORTY, 2018).

METODOLOGIA

Essa pesquisa foi aprovada pelo Programa Institucional de Bolsas de Iniciação Científica – PIBIC – Univértix. Este trabalho foi um estudo descritivo experimental com abordagem quali quantitativa, buscando utilizar *softwares* que

realizam cálculos matemáticos e análise por inteligência artificial (AI), visando a o estudo de moléculas bioativas com potencial farmacológico (CATHALA, MOORLEY, 2018). As etapas que compreendem esse estudo foram: busca pelo ligante, alinhamento do ligante, análise filogenética e análise físico-química *in silico*.

BUSCA PELO LIGANTE

Para que seja feita uma análise de compostos contra o novo coronavírus é necessário a busca pelo *Receptor Binding Domain* (Domínio de Ligação ao Receptor). O RBD, fornece o local exato da interação da proteína *spike* presente na variante ômicron (LAN *et al.*, 2020). Essa análise foi realizada através do *Protein Data Bank* (<https://www.rcsb.org/>) obtendo a proteína complexada com algum composto neutralizante, este forneceu informações sobre o ligante da proteína *spike*. Essa foi uma estratégia utilizada através de uma adaptação de outros trabalhos semelhantes sobre proteínas (SOARES, 2015).

ALINHAMENTO DO LIGANTE

Para obtenção de compostos semelhantes ao ligante foi realizada a busca no banco de dados *Uniprot (Universal Protein)* (Proteínas Universais). Nesse, se realizou a busca por proteínas que apresentavam domínios funcionais com identidade favorável ao ligante (MORGAT *et al.*, 2020).

Os cálculos de alinhamento foram feitos a partir do *software BLAST* (<https://www.uniprot.org/blast/>) que fornecerá alinhamento de proteínas semelhantes depositadas anteriormente no *Swiss Prot* (Proteínas Suíças) (<https://www.sib.swiss/swiss-prot>). No alinhamento, a sequência proteica em estudo foi confrontada com estruturas validadas (SOARES, 2015). Posteriormente foi realizado um segundo alinhamento. Esse processou a sequência em estudo através do *software MEGA11 (Molecular Evolutionary Genetics Analysis Version 11)* (<https://www.megasoftware.net/>) possibilitando a retirada de peptídeo de sinal, caso o mesmo esteja presente (TAMURA, STECHER, KUMAR, 2021).

ANÁLISE FILOGENÉTICA

Essa análise foi necessária com o intuito de predizer possíveis ancestrais comuns e relações de ancestralidade com outros vírus e organismos. Nessa etapa, foi utilizado o *software MEGA11*, capaz de processar sequências em formato *FASTA* (formato padrão de armazenamento de sequências de DNA, RNA e proteínas). Assim, foi possível realizar diferentes análises gerando uma topologia da proteína em estudo (SOARES, 2015).

ANÁLISE FÍSICO-QUÍMICA IN SILICO

O estudo de parâmetros físico-químicos, *in silico*, foi imprescindível para realização de posteriores testes *in vitro*, servindo como um funil para economia de tempo e recursos. Os melhores parâmetros para predição *in silico* são: índice de estabilidade e *Grand average of hydropathicity (GRAVY)*, utilizando o *software ProtParam (Protein Parameters) ExPasy (Expert Protein Analysis System)* (<https://web.expasy.org/protparam/>) (DUTTA, BANERJEE, CHAKRABORTY, 2018).

RESULTADOS E DISCUSSÕES

Através da busca no banco de dados do *Protein Data Bank* foi encontrado uma estrutura da proteína spike complexada com anticorpo neutralizante com resolução de 1,78Å (PDB DOI: 10.2210/pdb7JMP/pdb) que apresentou o ligante com a nomenclatura: “2-acetamido-2-deoxy-beta-D-glucopyranose”, com fórmula molecular C₈ H₁₅ O₁₆ (WU *et al.*, 2020).

A busca por proteínas com domínio do ligante foi realizada no *Uniprot*, o que resultou em 1 composto. Para realização de alinhamento buscando domínios semelhantes foi utilizado o *BLAST*. Em busca do refinamento dos resultados, a proteína identificada foi confrontada com as sequências proteicas depositadas no *Swiss Prot*, resultando em uma homologia com 36 sequências (SOARES, 2015). Como parâmetro de qualidade, foram utilizados somente as sequências que retornaram com valores acima de 40% de identidade, esse filtro resultou em 22 sequências de proteínas (LIBERMAN, 2004). As sequências foram consultadas,

através dos seus códigos de depósito individual no *Uniprot*. Em seguida, foi retirado o peptídeo sinal presentes localizados em domínios extracelulares utilizando o *software MEGA11* (TAMURA, STECHER, KUMAR, 2021).

Tendo em vista as sequências alinhadas foi realizada a análise filogenética a partir do *software MEGA11*, pelo método *Construct/Test Neighbor-Joining Tree* resultando em linhagens de diferentes espécies, com ancestrais comuns, pertencentes a mesma família de peptídeoglicanos (SOARES, 2015; TAMURA, STECHER, KUMAR, 2021).

Para a análise físico-química foram adicionadas as sequências *FASTA* de todas as 22 proteínas encontradas. Nessas, foi demonstrado que todas possuíam valor de *GRAVY* negativo (o que demonstrava caráter hidrofílico). Assim foi possível demonstrar que os compostos identificados nesse estudo, são solúveis em água e que sua administração não levará a precipitação (HAQ *et al.*, 2021). A avaliação do índice de estabilidade resultou na seleção de 8 compostos, uma vez que, estes se mostraram estáveis por apresentarem índices menores que 40. Esse índice é importante pois determina a estabilidade de compostos *in vitro* e posteriormente *in vivo* (HODA, TAJAJ, SALLAKU, 2021). Os dados da análise físico-química *in silico* foram compilados e analisados (TABELA 1).

Tabela 1: Dados físico-químicos obtidos no *software ProtParam tool - ExPASy*.

CÓDIGO DE ADESÃO	GRAVY	ÍNDICE DE INSTABILIDADE
Q96LB9.1	0.070	46.21
A1A547.2	-0.016	42.86
Q96LB8.2	-0.088	39.18
QOVBO7.1	-0.146	49.84
O75594.1	-0.262	59.51
B5T255.1	-0.271	42.06
Q8SPP7.1	-0.261	40.00
O888593.1	-0.258	58.33
Q9GK12.1	-0.189	42.88
Q9JLN4.1	-0.267	51.66
Q9VYX7.1	-0.028	36.09
O76537.1	-0.147	30.30
Q70PY2.2	-0.438	44.10
Q765P3.1	-0.332	38.80
Q70PR8.1	-0.250	34.06
Q9VS97.1	-0.279	31.51
Q9VXN9.1	-0.635	60.69

Q765P4.1	-0.309	39.76
Q8VCS0.1	-0100	38.42
Q866Y3.1	-0.156	43.24
Q96PD5.1	-0.129	45.54
Q8SXQ7.1	-0.223	48.38
TOTAL	-----	8

Fonte: ProtParam tool – ExPASy.

CONSIDERAÇÕES FINAIS

Existe uma grande variedade de compostos que apresentam identidade ao ligante da proteína *spike* (espigão) da variante ômicron. Observando os resultados deste estudo foi identificado que a proteína possui domínios funcionais exploráveis como alvo de novos tratamentos. Esses podem seguir como uma estratégia de alvo o domínio de ligação ao receptor da *spike*, visto que, sem essa proteína o vírus não consegue internalizar na célula hospedeira. Importante salientar que os estudos feitos a partir da análise físico-química *in silico* identificaram 8 moléculas hidrossolúveis. Portanto, é possível concluir que esta pesquisa contribuiu para diminuição de gasto, tempo e que esses compostos apresentam características físico-químicas que os tornam potenciais candidatos a novos fármacos, sendo necessário estudos moleculares, mesmo que *in silico*, que levem à caracterização estrutural das 8 moléculas identificadas, avançando assim em mais uma etapa na busca do desenvolvimento de compostos bioativos que combatam esse vírus.

REFERÊNCIAS

BARRETO-VIEIRA, D. *et al.* SARS-CoV-2: Ultrastructural Characterization of Morphogenesis in an In Vitro System. **Viruses**. Rio de Janeiro, Brasil, v.14, n. 2, p.0. Janeiro de 2022.

CATHALA, X.; MOORLEY, C. How to appraise quantitative research. *Evid Based Nurs. Evidence-Based Nursing*. Londres, Reino Unido, v. 21, n. 4, p. 99-101. Outubro de 2018.

DUTTA, B.; BANERJEE, A.; CHAKRABORTY, P.; BANDOPADHYAY, R. In silico studies on bacterial xylanase enzyme: Structural and functional insight. **Journal of Genetic Engineering & Biotechnology**. Bengala Ocidental, Índia, v.16, n. 2, p. 749–756. Dezembro de 2018.

GASTEIGER, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. **The Proteomics Protocols Handbook**. Suíça, v.1, n. 1, p. 571-607. Outubro de 2005.

GURUPRASAD, K.; REDDY, B.; PANDIT, M. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. **Protein Engineering**. Suíça, v.4, n. 2, p. 155-161. Dezembro de 1990.

HAQ, A. *et al.* Annotation of Potential Vaccine Targets and Design of a Multi-Epitope Subunit Vaccine against *Yersinia pestis* through Reverse Vaccinology and Validation through an Agent-Based Modeling Approach. **Vaccines**. Suete, Paquistão, v.9, n. 11, p. 1327-1346. Novembro de 2021.

HE, X.; HONG, W.; PAN, X.; LU, G.; WEI, X. SARS-CoV-2 Omicron variant: Characteristics and prevention. **MedComm**. Chengdu, China, v.2, n. 4, p. 838-845. Dezembro de 2021.

HODA, A.; TAJAJ, M.; SALLAKU, E. In silico Structural, Functional and Phylogenetic Analyses of cellulase from *Ruminococcus albus*. **Journal of Genetic Engineering & Biotechnology**. Tirana, Albania, v. 19, n. 1, p. 58-73. Abril de 2021.

KAUR, T.; MADGULKAR, A.; BHALEKAR, M.; ASGAONKAR, K. Molecular Docking in Formulation and Development. **Drug Discovery Technologies**. Maharashtra, Índia, v. 19, n. 1, p. 30-39. Junho de 2019.

KHAN, M. *et al.* COVID-19: A Global Challenge with Old History, Epidemiology and Progress So Far. **Molecules**. Basileia, Suíça, v. 26, n. 1, p. 39-64. Dezembro de 2020.

ALKHATIB, M. *et al.* Update on SARS-CoV-2 Omicron Variant of Concern and Its Peculiar Mutational Profile. **Microbiology Spectrum**. Roma, Itália, v.10, n. 2, p. 2732-2749. Março de 2022.

KYTE, J.; DOOLITTLE, F. A simple method for displaying the hydropathic character of a protein. **Molecular Biology**. Califórnia, Estados Unidos, v. 157, n. 1, p. 105-132. Maio de 1982.

LAN, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. **Nature**. Pequim, China, v.581, n. 7807, p. 215-220. Março de 2020.

LIBERMAN, F. (2004). **Análise dos fatores determinantes para a qualidade da anotação genômica automática**. Tese de Mestrado (Biotecnologia e Ciências genômicas) – Universidade Católica de Brasil. Brasília, 2004.

MORGAT, A. *et al* Enzyme annotation in UniProtKB using Rhea. **Bioinformatics**. Genebra, Suíça, v.36, n. 6, p.1896–1901. Março de 2020.

SOARES, A. C. (2015). **Identificação de lipocalinas com domínio de ligação a histamina e serotonina e validação de genes de referência em transcriptomas de glândula salivar, intestino e ovário do carrapato *Amblyomma sculptum***. Tese de doutorado (tese em Bioquímica Aplicada) - UFV, Departamento de Bioquímica e Biologia Molecular. Viçosa. 113p.

TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: Molecular Evolutionary Genetics Analysis Versão 11. **Molecular Biology and Evolution**. Pensilvânia, Estados Unidos, v.38, n. 7, p. 3022-3027. Junho de 2021.

TANTIVITAYAKUL, P. *et al*. Identification and *in silico* functional prediction of lineage-specific SNPs distributed in DosR-related proteins and resuscitation-promoting factor proteins of *Mycobacterium tuberculosis*. **Heliyon**. Bangkok, Tailândia, v. 6, n.12, p. e05744, 2020.

WU, N. *et al*. Um modo de ligação alternativo de anticorpos IGHV3-53 ao domínio de ligação do receptor SARS-CoV-2. **Cell Reports**. Califórnia, Estados Unidos, v.33, n. 3, p. 108274, 2020.