

Título do Projeto de Pesquisa e Produtividade Aprovado: Análise in silico da proteína S do coronavírus Sars-CoV-2

Autor do Projeto: Luciana Maria de Hollanda

Área/Subárea Capes: Saúde

O primeiro caso da doença coronavírus 2019 (COVID-2019) foi identificada na cidade de Wuhan na china em dezembro de 2019. Até hoje, este vírus rapidamente se espalhou em 188 países e territórios, infectando mais de 244 milhões de pessoas com 4,95 milhões de mortes em todo o globo terrestre.

O vírus Sars-CoV-2 é um vírus do gênero *Betacoronavirus* da família *Coronaviridae*. Este gênero possui envelope e fita de RNA de sentido positivo, que infecta tanto humanos como animais. O genoma codifica proteínas estruturais, acessórias e não estruturais. Entre as proteínas acessórias, as proteínas estruturais são codificadas nas regiões “downstream” ao genoma. Estas incluem a proteína de espícula (S), a proteína do envelope (E), a proteína de membrana (M) e a proteína do nucleocapsídeo ligado ao RNA (N).

O coronavírus infecta as células humanas utilizando a glicoproteína S que se liga ao receptor de transmembrana angiotensina conversor da enzima 2 (ACE2) que se localiza no epitélio pulmonar e, no epitélio do intestino delgado em humanos. Diversos estudos bioquímicos relataram a alta homologia do receptor. Quando comparado com o SARS-CoV, a proteína S na posição 455 (leucina) e 486 (fenilalanina) no SARS-CoV-2 mostrou aumento de interação com o ACE2.

Além disso, a proteína S, foi identificada como o principal antígeno imunodominantes deste vírus. A avaliação de pacientes com SARS-CoV-2 revelou que anticorpos de ligação e neutralização visam principalmente o domínio de ligação ao receptor da subunidade S.

Então vacinas construídas e direcionadas para esta proteína tem grandes chances de êxito. Até o presente momento devido ao grande fluxo de trabalho por parte de diversas agências de pesquisa ao redor do mundo, foram geradas muitas sequencias desta proteína, pertencente a este vírus. De acordo com dados

depositados no Banco de Dados do NCBI-SARS-CoV-2, tem-se sequenciado 2.065.755 nucleotídeos e 982.016.320 sequências de aminoácidos relacionados a proteína S.

Para processarmos estes dados, torna-se necessário a utilização de ferramentas robustas como por exemplo, a bioinformática.

A bioinformática é uma das disciplinas que mais evoluem com o tempo. Resumidamente é a aplicação de ferramentas e técnicas *in silico* para a gestão e análise de dados biológicos. Existem várias ferramentas disponíveis tanto na versão paga, como na versão gratuita para manipular e analisar estes dados. Entre estas ferramentas encontramos o biopython.

Biopython é um conjunto de ferramentas disponíveis gratuitamente para computação biológica, escrito em Python, por uma equipe internacional de desenvolvedores. O código fonte é disponibilizado sob a licença Biopython, que é extremamente liberal e compatível com quase todas as licenças do mundo.

Desta forma, através da análise com o Biopython, torna-se possível prever a forma e a função da proteína S, a partir da sequência primária de aminoácidos desta proteína e, ainda se torna possível determinar áreas na estrutura da proteína, onde as moléculas ou fármaco podem se ligar, evitando que o vírus penetre na célula.

Assim percebe-se que a bioinformática é essencial para agilizar o processo de investigação da proteína S do Sars-CoV-2.

