

MUSTER - SISTEMA DE MINERAÇÃO DE CONTATOS EM COMPLEXOS PROTEÍNA-LIGANTES: SELEÇÃO E INTERFACE

Wellington G. Silva¹ (IC), Carlos H. da Silveira² (PQ), João Paulo R. Romanelli³ (PQ)
^{1,2,3}Universidade Federal de Itajubá – campus Itabira.

Palavras-chave: Bioinformática. COVID-19. EasyVS. Mineração de Dados. Triagem Virtual.

Introdução

A temática desta pesquisa de iniciação científica tem como base o projeto financiado pela Vale S.A através do convênio Unifei-Vale de Iniciação em Engenharia, com o título “Previsão *in silico* de candidatos a fármacos antivirais como terapêutica auxiliar da COVID-19”.

Na natureza, é comum que se encontre dados na ordem de grandeza dos milhões ou bilhões de registros. No processo de descoberta de fármacos (*drug discovery*), não é diferente. Nesse sentido, a Triagem virtual de ligantes, do inglês *Virtual Screening*, é indispensável para auxiliar na descoberta de medicamentos por intermédio da pesquisa de numerosas bibliotecas de pequenas moléculas. Essa técnica tem como objetivo identificar estruturas com maior probabilidade de se ligar a um determinado alvo farmacológico, geralmente um receptor ou enzima.

Nesse escopo, a Triagem Virtual auxilia na seleção de moléculas num processo *in silico*, isto é, via simulação computacional, uma vez que, com o uso de *softwares* adequados é possível estimar a interação entre proteínas e possíveis ligantes, diminuindo assim custos de pesquisa nas fases *in vitro* e *in vivo*.

Tendo em vista a pandemia causada pelo vírus SARS-Cov-2 e a necessidade urgente de se encontrar um fármaco com ação inibidora para atuar paralelamente às vacinas que surgissem no decorrer do processo, a aplicação de conceitos do ramo da bioinformática, química, biologia e computação na busca pelo ligante com afinidade requerida foram as motivações desta pesquisa.

Metodologia

Durante o processo de triagem virtual, utiliza-se de ferramentas de ancoragem molecular (*molecular docking*), e nesta pesquisa utilizou-se de forma constante a plataforma web chamada EasyVS[4] desenvolvida pelo nosso grupo de pesquisa INSILICO. Para verificar a acurácia desta ferramenta, realizou-se testes de *redocking*, cujo papel é verificar um sítio ativo com ligante e posição de ligação já conhecidos e resolvidos por métodos experimentais da literatura, repetindo o atracamento na região da proteína utilizando o mesmo alvo e mesmo ligante, esperando-se obter o *docking* com RMSD (*Root Mean Square Deviation*) baixo, indicando

a boa acurácia desejada.

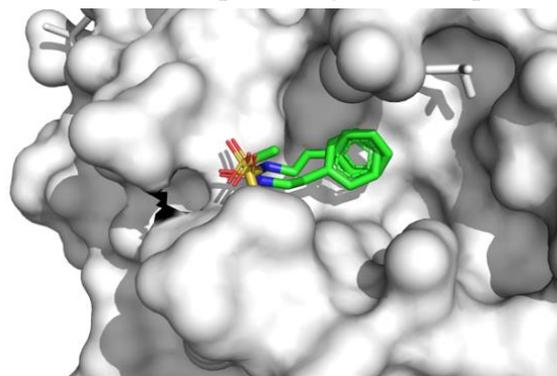
O Protein Data Bank (PDB) é uma plataforma internacional contendo milhões de dados de estruturas de biomoléculas resolvidas experimentalmente e armazenados na forma de dados possíveis de serem processados por métodos computacionais para análise.

Nesse contexto, utilizando a API de dados do PDB, realizou-se a mineração de dados (.pdbs) relacionados à *main protease* (Mpro) do vírus *Sars-Cov-2*, responsável pela pandemia de Covid-19, através do acesso à seção de recursos COVID-19/SARS-CoV-2.

Utilizando um *script*, temporariamente chamado de “musteR”, que realiza a mineração de padrões de *docking* de ligantes num alvo, por análise de agrupamentos de densidade de contatos e empacotamentos atômicos, está se desenvolvendo uma *interface web* em linguagem R através do *framework Shiny*, o qual é um pacote que facilita a criação e desenvolvimento de *web apps* interativos, para que se faça a integração simplificada entre usuário e código.

Resultados e discussão

Rotinas de testes de *redocking* foram feitos no EasyVS, obtendo-se resultados satisfatórios como se pode verificar no exemplo da Figura 1, em que foi feito



redocking do ligante cristalográfico de PDB ID JFM com o alvo de PDB ID 5R7Y, obtendo RMSD de 1,431 Å. Figura 1 – *Redocking* no EasyVS. Visualização no *software Pymol*[3].

Buscando preparar uma base de dados de pdbs da Mpro e seus respectivos ligantes cristalográficos, através do método de obtenção de dados API do PDB, utilizou-se o filtro do RCSB PDB considerando a Mpro como

alvo. Dessa forma, encontrou-se inicialmente cerca de 500 pDBs ID's da Mpro, cuja lista foi comparada, através de *scripts* em Python, com 107 PDB's de um artigo [2] do início de 2020 publicado pelo INSILICO, inferindo-se que os 107 alvos estavam na base de PDB's atualizada, ou seja, não houveram exclusões mas, sim, 393 novas adições de PDB's devido aos avanços das pesquisas acerca da COVID-19.

A interface do musteR está em desenvolvimento e sua versão inicial básica está evidenciada na Figura 2.

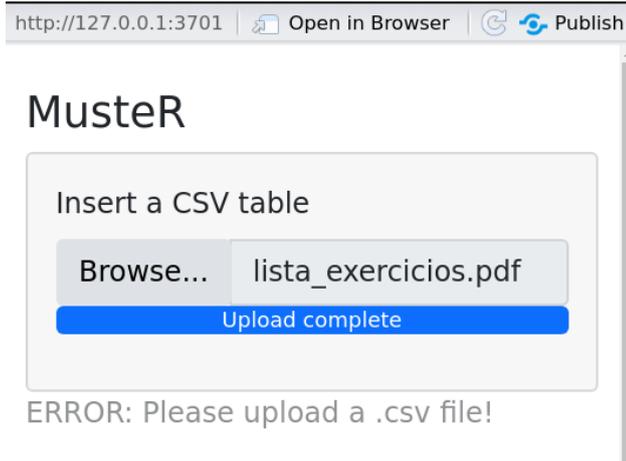


Figura 2 – Ideia básica inicial da interface do musteR.

Espera-se, ainda, utilizar a linguagem de marcação Markdown visando integrar código e documentação.

Nessa tela da Figura 2, o usuário deverá fornecer uma tabela em csv contendo os pDBs que deseja utilizar no *script*.

O objetivo é, após o usuário fazer o *upload* e o *script* rodar, obter-se como saída um gráfico tridimensional interativo que exiba as interações por meio de nuvem de *tags*, contendo os aminoácidos, cuja interação alvo-ligante está atrelada ao tamanho da *tag*, ou seja, quanto maior a *tag*, mais interações o aminoácido do ligante em questão faz com o alvo. Tal gráfico pode ser verificado na Figura 3 para os PDBS 7NF5 (Crystal structure of MG-132 covalently bound to the main protease (3CLpro/Mpro) of SARS-CoV-2 in spacegroup C2) e 6XQU (Room-temperature X-ray Crystal structure of SARS-CoV-2 main protease in complex with Boceprevir).

Nessa Figura 3, observa-se que a *tag* do aminoácido Q189 é maior para o alvo 7NF5 em comparação ao alvo 6XQU, evidenciando que Q189 tem mais influência no alvo 7NF5.

O musteR também permite a sobreposição de vários ligantes no mesmo *pocket* (sítio ativo), permitindo, assim, análises estatísticas de interações para vários ligantes no mesmo alvo.

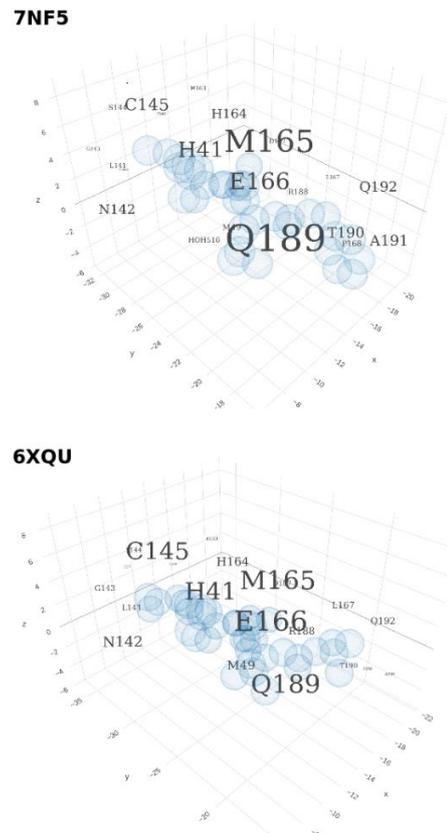


Figura 3 – Gráfico esperado a ser apresentado na interface.

Conclusões

Pode-se concluir que o musteR pode ser uma ferramenta muito útil na análise estatística de interações em complexos alvo-ligante, uma vez que pode fornecer de forma interativa e gráfica padrões de interações.

Agradecimento

Gostaria de agradecer à oportunidade da bolsa, financiada pela Vale S.A.

Referências

- [1] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
- [2] Rafael E. O. Rocha, Fabiana C. Guedes, Carlos H. da Silveira, *et al.* (2021). A higher flexibility at the SARS-CoV-2 main protease active site compared to SARS-CoV and its potentialities for new inhibitor virtual screening targeting multi-conformers, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2021.1924271
- [3] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC
- [4] VELOSO, W. N. P. EasyVS: uma ferramenta para triagem virtual mista baseada em alvo e ligante. Tese de Doutorado - Universidade Federal de Minas Gerais. Belo Horizonte. 2019. Disponível em: <<http://hdl.handle.net/1843/30754>>, acesso em 12 nov. 2021.