

# MusteR - Sistema de Mineração de Contatos em Complexos Proteína-Ligantes

Wellington Gonçalves Silva<sup>1</sup>, Cláudio Guimarães Matos Júnior<sup>1</sup>, Wandré Nunes de Pinho Veloso<sup>1</sup>, João Paulo Roquim Romanelli<sup>1</sup>, Carlos Henrique da Silveira<sup>1</sup>

<sup>1</sup>Instituto de Ciências Tecnológicas -- Universidade Federal de Itajubá - UNIFEI, Itabira, MG. Brasil.

{welington.kaaminsk2002, claudiogmj, wandreveloso, joaoromanelli, carlos.silveira}@unifei.edu.br

**Abstract.** *The COVID-19 pandemic has shown how high the demand is for computational tools that aid in the process of discovering new drugs. In this context, analyzing how ligands interact with a biomolecular target has proven valuable in identifying compounds with greater pharmacological potential. To this end, we propose MusteR, a web tool for mining contact pattern motifs between protein-ligand complexes, based on atomic packing density. Its use will be demonstrated in characterizing the complex between the drug Boceprevir and the Mpro protease, an important therapeutic target in the search for antivirals against SARS-CoV-2.*

**Resumo.** *A pandemia de COVID-19 mostrou quão alta é a demanda por ferramentas computacionais que auxiliem no processo de descoberta de novos fármacos. Nesse contexto, a análise de como ligantes se interagem com um alvo biomolecular tem se mostrado valiosa na indicação de compostos com maior potencial farmacológico. Para tanto, propõe-se MusteR, uma ferramenta web de mineração de padrões de contatos entre complexos proteína-ligante, pela densidade de empacotamentos atômicos. Será demonstrado seu uso na caracterização do complexo entre o fármaco Boceprevir e a protease Mpro, um importante alvo terapêutico na pesquisa por antivirais contra o SARS-CoV-2.*

## 1. Introdução

A busca por novos fármacos é uma tarefa complexa e desafiadora que envolve a identificação de compostos químicos ou ligantes capazes de interagir com alvos biológicos específicos e promover efeitos terapêuticos desejados. Nesse contexto, a análise *in silico* de como esses ligantes se acomodam quimicamente em seu alvo molecular tem se mostrado uma técnica valiosa para identificar compostos com maior potencial farmacológico. No entanto, a interpretação dos resultados pode ser difícil, especialmente quando se trata de identificar os padrões de interação entre ligantes e alvos [Shaker et al, 2021].

Nesse contexto, torna-se relevante o MusteR - uma ferramenta de mineração de padrões de contatos entre complexos proteína-ligante, por análise de agrupamentos por densidade de empacotamentos atômicos. Através de uma interface *web* baseada em linguagem R e no *framework* Shiny [Chang et al. (2023)], o MusteR facilita a

identificação de contatos com maior interação e oferece uma abordagem mais intuitiva para a análise dos resultados da ancoragem molecular, que consiste num posicionamento energética e estruturalmente aceitável do ligante sobre o alvo, geralmente sobre uma região invaginada da superfície proteica conhecida como bolsão ou *pocket*. Com isso, o MusteR pode ser uma ferramenta auxiliar interessante para pesquisadores que buscam desenvolver novos fármacos e contribuir para avanços significativos no campo da saúde.

MusteR infere a densidade de empacotamentos calculando as áreas de contatos atômicos. Para isso, é montada uma rede de contatos representado por um grafo k-partido, não direcionado, em que os nós representam átomos e as arestas as áreas de contatos entre esses átomos. O grafo é k-partido porque somente são permitidas arestas entre átomos que não sejam da mesma molécula, como exemplificado na Figura 1 [Araujo, 2019b].

O uso de área de contato como pesos para as arestas do grafo se justifica por conta da correlação que existe entre essas áreas e as energias de ligação e outros parâmetros termodinâmicos do complexo ligante-alvo, como capacidade calorífica, entalpia e entropia [Baker & Murphy, 1997]. O que enfatiza a importância que tem o padrão químico e geométrico das superfícies de uma proteína para o desempenho de suas funções. As áreas de contatos são calculadas conforme metodologia descrita em [Araujo, 2019b].

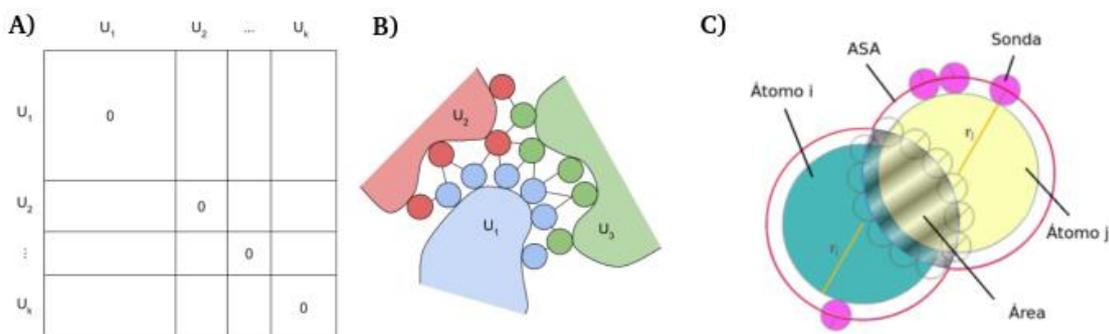


Figura 1. Descrição esquemática dos grafos k-partidos e da área de contato. A) Como o grafo k-partido é representado numa matriz de adjacências. B) Grafo 3-partido entre 3 biomoléculas diferentes ( $U_i$ ), tendo átomos da superfície como nós. C) Área de contato entre átomos  $i$  e  $j$ , como a área não acessível por uma molécula de água ou sonda. (Fonte: Adaptado de [Araujo, 2019a]).

## 2. Arquitetura

O MusteR é uma ferramenta cliente/servidor que integra a facilidade de uso de uma interface gráfica *web* com a linguagem R, por meio do *framework* Shiny. Graças à biblioteca Shiny e suas funções, é possível proporcionar uma experiência interativa e intuitiva para os usuários no lado *frontend*. A interface *web* é acessada por meio do navegador, tornando os resultados facilmente visualizáveis. Todos os *workflows* e *scripts* necessários para a execução da análise são hospedados no lado *backend* da ferramenta (Figura 2).

### 2.1. Backend

O MusteR faz uso do Shiny *Server* para hospedar o aplicativo. O Shiny *Server* é um programa de *backend* de código aberto que cria um servidor *web* especificamente projetado para hospedar aplicativos Shiny. Com o Shiny *Server*, é possível hospedar

aplicativos em um ambiente controlado. Também é possível usar o *Shiny Server* para disponibilizar aplicativos pela Internet quando desejado. O *Shiny Server* hospeda cada aplicativo em seu próprio endereço da *web* e inicia automaticamente o aplicativo quando um usuário visita o endereço. Quando o usuário sai, o *Shiny Server* interrompe automaticamente o aplicativo.

## 2.2. Frontend

O propósito do *frontend* consiste em desenvolver uma interface gráfica de fácil compreensão para o usuário e que organize o fluxo de trabalho de maneira eficiente. Além disso, ele possui um papel fundamental no preenchimento exato das informações que serão inseridas no *backend*. A principal biblioteca empregada na construção do *MusteR*, denominada *Shiny*, proporciona a fusão do código R com outras linguagens, como o JavaScript, por meio da biblioteca *htmlwidgets*, que viabiliza a união de widgets HTML/CSS/JavaScript. Dessa forma, o *MusteR* faz uso do *framework* *Shiny* tanto como uma biblioteca visual no *frontend* quanto como servidor *backend web*.

No *frontend*, a aplicação *Shiny* é contida em um único script chamado *app.R* que possui três componentes: objeto *User Interface* (UI), a função *Server* e a chamada padrão do programa, *ShinyApp*. A UI controla o layout e a aparência do aplicativo. A função *server* contém as instruções para construir o aplicativo quando o usuário interage com o programa. Por fim, *ShinyApp* cria objetos *Shiny* a partir de um par explícito *UI/Server*.

Utilizou-se do pacote *R3dmol* [Su W, Johnston B., 2023] para a visualização molecular 3D, que emprega o *3Dmol.js* como *htmlwidgets*. O *3Dmol.js* [Rego, N., & Koes, D., 2015] é uma biblioteca JavaScript acelerada por *hardware*, orientada a objeto e projetada para permitir que usuários e desenvolvedores da *web* interajam com dados moleculares em navegadores modernos de *desktops* e dispositivos móveis com desempenho semelhante ao nativo. Seu foco é fornecer uma API completa para visualização de alta performance de moléculas online, que pode ser integrada a outros aplicativos da *web* para análise e quimioinformática. A biblioteca suporta uma variedade de estilos comuns e vários formatos de arquivo, incluindo *pdb*, *sdf*, *mol2*, *xyz* e *cube*. O *3Dmol.js* é utilizado para visualizar dados moleculares por desenvolvedores de aplicativos da *web*, criadores de HTML e usuários finais. O *htmlwidgets* é uma biblioteca de visualização em JavaScript que permite a criação de novos *widgets* HTML usando uma estrutura que une o R e o JavaScript. Ele possibilita a produção de visualizações interativas em documentos R *Markdown* e aplicativos da *web* *Shiny*.

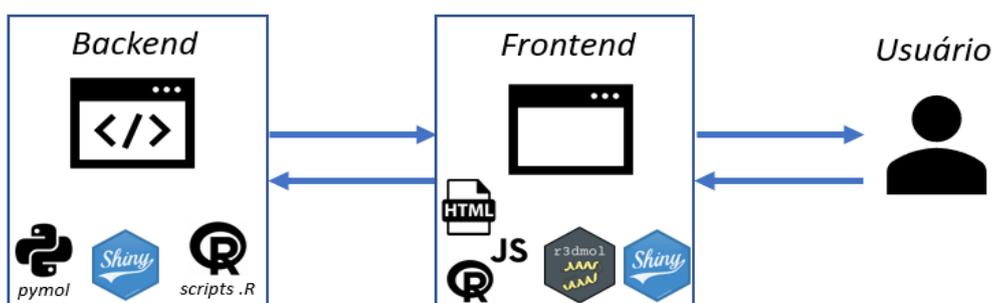


Figura 2. Arquitetura *MusteR*. (Fonte: autores).

### 3. Funcionalidades

#### 1. Página: Home - download ou upload do arquivo PDB:

Na página inicial do MusteR, o usuário tem a opção de fazer o *download* da molécula diretamente do site RCSB PDB (<https://www.rcsb.org>) ou fazer o *upload* de um arquivo local. Para isso, é necessário preencher as informações solicitadas, como o código PDB, o código e o ID do ligante, na linha indicada.

#### 2. Página: Interactive Map – visualização das interações alvo-ligante mais importantes:

Após o usuário clicar em um dos botões de submissão, os scripts do *backend* iniciam o processamento e, quando concluído, são geradas duas visualizações em uma nova página que é chamada automaticamente. À esquerda, tem-se a visualização molecular tridimensional do alvo com o ligante e, à direita, um mapa de *tags* de resíduos codificados em 3 letras e o id. Ambos os gráficos são atualizados quando o botão deslizante é movido pelo usuário, exibindo os resíduos do alvo e do ligante com base em sua área, indo das duas maiores até as sete maiores áreas de contatos.

##### 2.1. Visualização molecular tridimensional interativa das interações alvo-ligante:

Nessa visualização 3D, é possível ver e mover o complexo ligante-alvo e visualizar os rótulos contendo os resíduos com os quais eles mais interagem. É possível também alterar a quantidade de resíduos exibidos com base na área alterando o valor do botão deslizante, como citado anteriormente.

##### 2.1.1. Botão de escolha de estilo de visualização molecular:

Dentro do escopo da visualização 3D, é possível selecionar um botão que alterna o estilo de visualização do alvo entre "*sticks*" (bastões) e "*spheres*" (esferas), facilitando a compreensão visual.

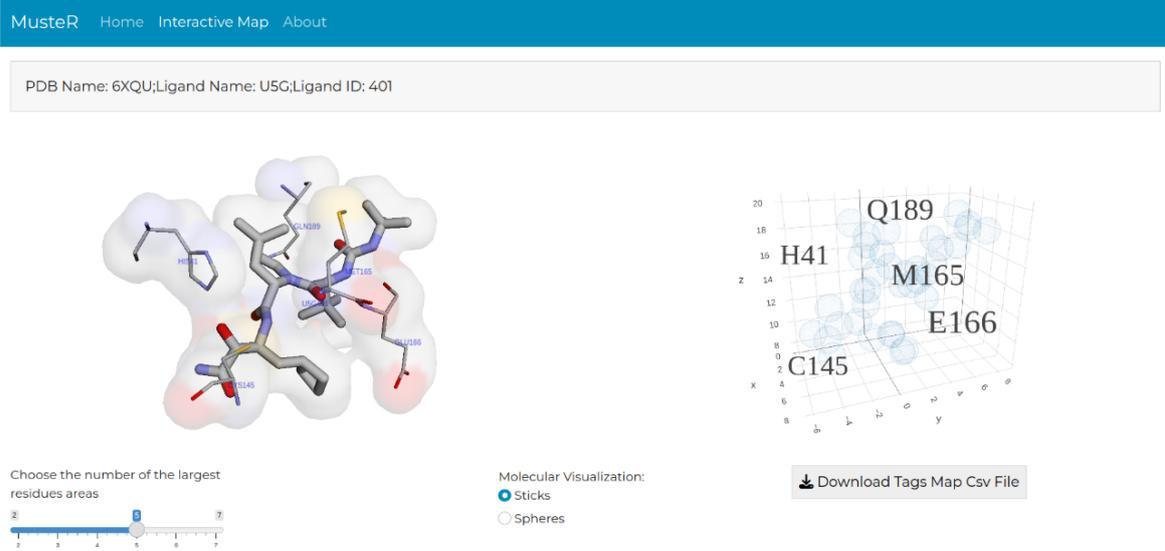
##### 2.2 Visualização via mapa de tags interativo com base nas maiores áreas selecionadas:

Nessa visualização de mapa de *tags* de resíduos, é possível visualizar de forma intuitiva os resíduos que mais interagem com o alvo. Isso ocorre porque o tamanho da *tag* (código do resíduo) exibido é proporcional à sua área de contato com o ligante. Ou seja, quanto maior a área de contato do resíduo do alvo com o ligante, maior será a sua exibição no mapa de *tags*. Essa apresentação visual é bastante útil para identificar quais resíduos do alvo estão mais bem empacotados com o ligante, o que implica também numa maior probabilidade de interações. Apresenta ainda, em cor azul, os átomos dos ligantes em mapa de pontos (*dots*).

##### 2.2.1 - Botão de download do mapa de tags:

No escopo da visualização via mapa de *tags*, é possível realizar o *download* do arquivo .csv contendo os dados do mapa de *tags*. Isso pode ser útil para o usuário compor análises e visualizações mais complexas, em outros aplicativos como o Pymol ou VMD.

Na Figura 3, pode-se visualizar a interface do MusteR após a submissão e o processamento dos dados para o alvo 6XQU.



**Figura 3.** Página após o processamento no *backend* para o PDB 6XQU. Percebe-se entre os 5 resíduos com maiores áreas de contato estão: E166, M165, Q189, H41 e C145. (Fonte: autores).

#### 4. Demonstração

Uma demonstração de uso do MusteR pode ser vista em <https://bit.ly/MusteR-demo> e seu acesso em <https://bit.ly/MusteR-app>. O vídeo mostra os principais resíduos que participam do complexo envolvendo a *main protease* (Mpro) do SARS-CoV-2 com um ligante inibidor Boceprevir. O SARS-CoV-2 foi o nome dado ao vírus responsável pela recente pandemia de COVID-19. A Mpro é uma das proteases que participam do processo de replicação viral. Logo, sua inibição faz dela um alvo estratégico para antivirais no tratamento da COVID-19. O Boceprevir é um fármaco já aprovado e em comercialização para tratamento da Hepatite C, por inibição de uma de suas proteases. A intenção era verificar se Boceprevir também poderia inibir a protease do SARS-CoV-2, processo conhecido como reposicionamento de fármaco [Kneller et al, 2020]. No vídeo de demonstração foi usado a estrutura resolvida da Mpro de código 6XQU, com o ligante Boceprevir de código U5G e id 401. A 6XQU pode ser buscada diretamente no repositório de biomoléculas RSCB (<https://www.rcsb.org>). Uma vez lida, o MusteR irá selecionar os resíduos com maiores áreas de contatos no entorno do Boceprevir. Percebe-se que entre os 5 resíduos com as maiores áreas de contato estão a HIS41 e a CYS145, já conhecidas por serem os principais resíduos no processo de catálise da Mpro [Rocha et al, 2021]. Logo, isso sugere que o Boceprevir teria chance de ser um forte inibidor dessa protease.

#### 5. Considerações Finais

MusteR é um sistema ainda em desenvolvimento e várias outras funcionalidades estão previstas. Algumas até já estão sendo codificadas como: oferecer mais opções de visualização ao usuário, incluindo tanto a parte molecular, como também dos grafos subjacentes; sincronizar automaticamente as visualizações entre o modelo molecular e a *tag* de resíduos; permitir visualizações agrupadas quando houver mais de um complexo ligante-alvo, como quando se testam vários candidatos a fármaco para um mesmo alvo; incluir algoritmos de mineração de dados em grafos, como agrupamentos de subgrafos

densamente conectados; melhorar a caracterização das regiões *hotspots* ou farmacofóricas, aquelas regiões da superfície do alvo mais propícias a um planejamento racional de candidatos a fármacos; incorporar também a visualização e análise de contatos para além de ligantes-alvos, como proteína-proteína, proteína-ácido nucleicos etc.

## 6. Agradecimentos

Um agradecimento especial à Vale S.A, pelo apoio a este projeto.

## 7. Referências

- Araujo, B. M., & da Silveira, C. H. (2019a). GAPIN: uma Ferramenta para Visualização e Análise de Redes de Interações Atômicas Intermoleculares [UFMG]. [https://e3sys.com.br/ppg21/admin/files/tese/document\\_8FVMltzmUgcW.PDF](https://e3sys.com.br/ppg21/admin/files/tese/document_8FVMltzmUgcW.PDF)
- Araújo, B. M., Coelho, A. L., Silveira, S. A., Romanelli, J. P. R., de Melo-Minardi, R. C., & da Silveira, C. H. (2019b). GAPIN: Grouped and Aligned Protein Interface Networks. In bioRxiv (p. 520833). <https://doi.org/https://doi.org/10.1101/520833>
- Baker, B. M., & Murphy, K. P. (1997). Dissecting the energetics of a protein-protein interaction: the binding of ovomucoid third domain to elastase. *Journal of Molecular Biology*, 268(2), 557–569. <https://doi.org/10.1006/jmbi.1997.0977>
- Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2023). shiny: Web Application Framework for R. R package version 1.7.4.9002, <https://shiny.rstudio.com/>
- Rego, N., & Koes, D. (2015). 3Dmol.js: Molecular visualization with WebGL. *Bioinformatics*, 31(8), 1322–1324. <https://doi.org/10.1093/bioinformatics/btu829>
- Rocha, R. E. O., Chaves, E. J. F., Fischer, P. H. C., Costa, L. S. C., Barden, G. I., Cruz, L. E. G., Guedes, F. C., Silveira, C. H. da, Scotti, M. T., Camargo, A. D., Machado, K. S., Werhli, A. V., Ferreira, R., Rocha, G., & De Lima, L. H. F. (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*, 0(0), 1–22. <https://doi.org/10.1080/07391102.2021.1924271>
- Shaker, B., Ahmad, S., Lee, J., Jung, C., & Na, D. (2021). In silico methods and tools for drug discovery. *Computers in Biology and Medicine*, 137(July), 104851. <https://doi.org/10.1016/j.combiomed.2021.104851>
- Su W, Johnston B (2023). r3dmol: Create Interactive 3D Visualizations of Molecular Data. R package version 0.2.0, <https://github.com/swsoyee/r3dmol>