

## Desenvolvimento de uma Ferramenta Computacional para Clusterização Molecular e Identificação do Máximo Subgrafo Comum

Wenderson Rodrigues Lopes<sup>1</sup>; Gabriel Teixeira Guerra<sup>2</sup>; Christiane Mariotini Moura Vasconcellos<sup>3</sup>, Juliana Lopes Rangel Fietto<sup>2</sup>; Raphael De Souza Vasconcellos<sup>2</sup>

1 - Departamento de Informática; 2 - Departamento de Bioquímica e Biologia Molecular; 3 - Departamento de Medicina e Enfermagem

ODS3 - SAÚDE E BEM-ESTAR

### Introdução

O avanço das técnicas de bioinformática tem revolucionado o processo de descoberta e desenvolvimento de novos fármacos, tornando possível analisar grandes volumes de dados e prever interações moleculares de forma mais rápida e econômica. Ferramentas computacionais como o RDKit oferecem recursos essenciais para simulações *in silico*, permitindo a análise de similaridade estrutural, clusterização de compostos e a identificação de padrões químicos relevantes. Nesse contexto, a clusterização molecular e a identificação do Máximo Subgrafo Comum (MCS) destacam-se como estratégias poderosas para revelar características estruturais compartilhadas entre moléculas, auxiliando no desenho racional de fármacos.

### Objetivos

Desenvolver um Jupyter Notebook interativo, utilizando a biblioteca RDKit em Python e o Google Colab, para:

- Realizar a clusterização molecular de compostos químicos com base em *fingerprints* e no algoritmo de Butina.
- Identificar o Máximo Subgrafo Comum (MCS) entre moléculas agrupadas.
- Fornecer uma interface clara e de fácil utilização, que permita a análise de similaridade estrutural e a seleção de moléculas promissoras para estudos posteriores.

### Metodologia

Com o banco de dados de moléculas da PkIDB, foi aplicado o algoritmo de Butina juntamente com o algoritmo de tanimoto para obter as matrizes de similaridade, gerando os *clusters* das moléculas:

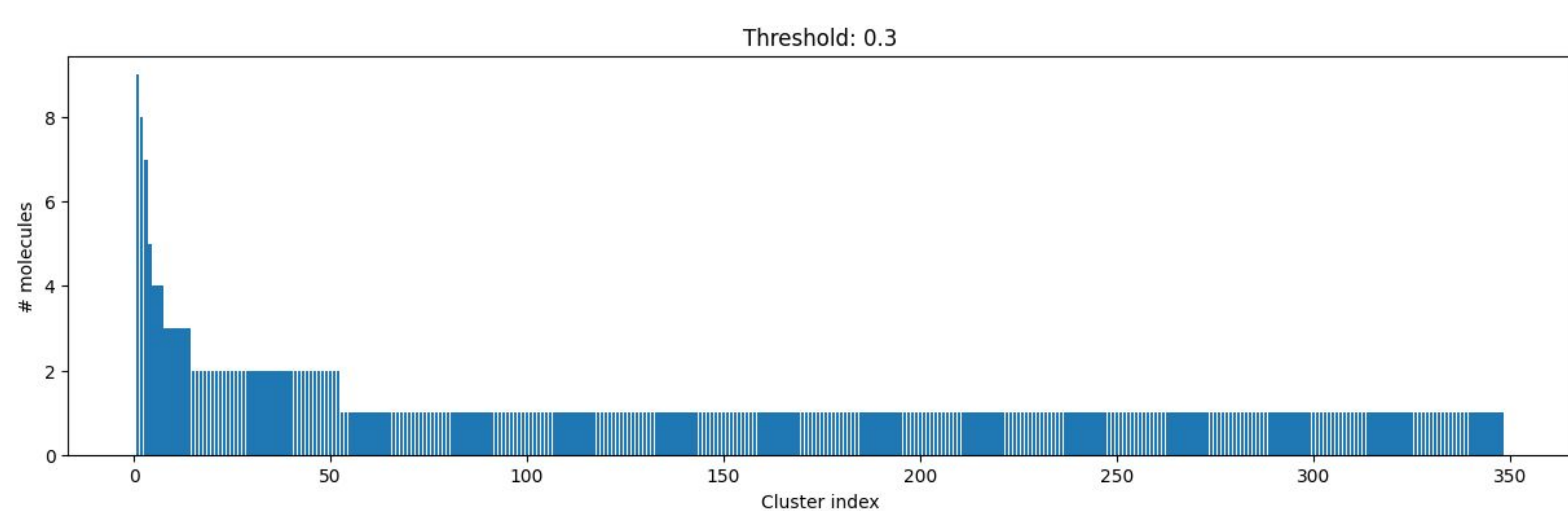


Gráfico 1 - Histograma indicando quantas moléculas existem em cada *cluster* gerado pela ferramenta.

O *cluster* gerado pela ferramenta teve como resultado a visualização das moléculas:

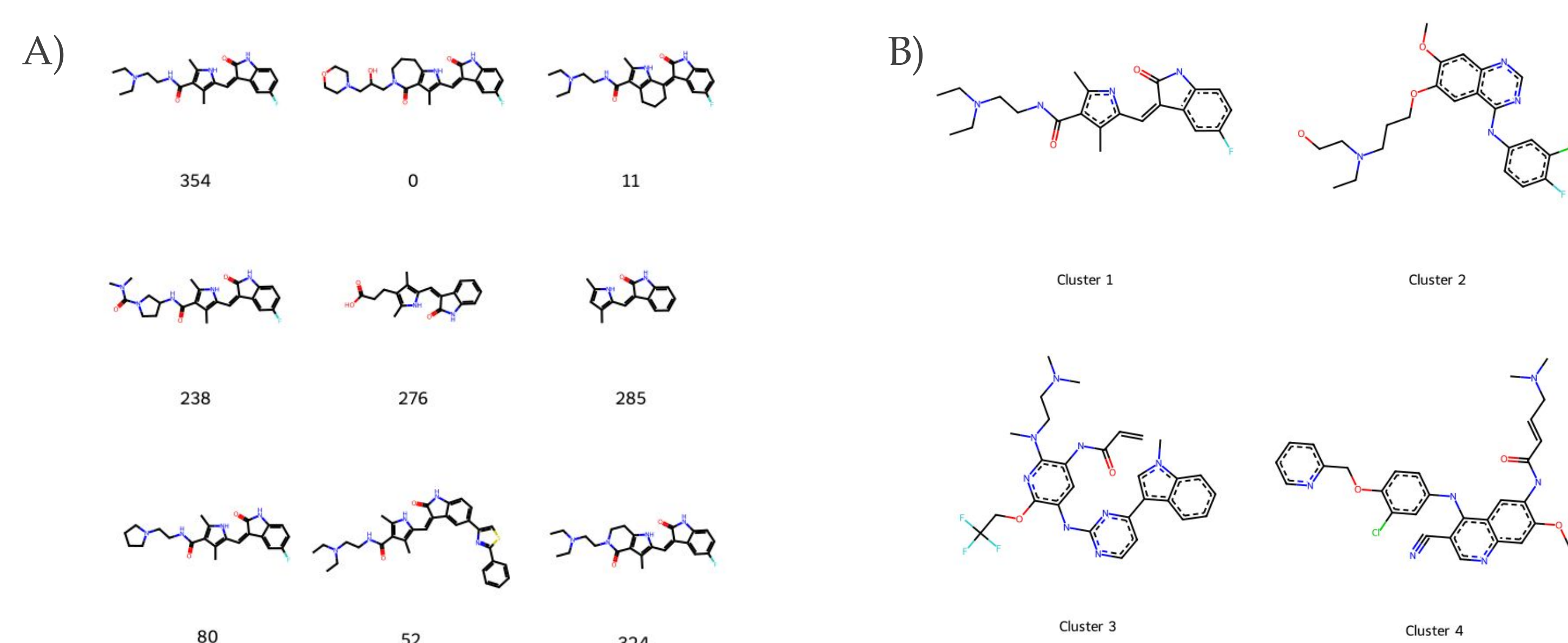


Figura 1: A - Moléculas presentes no maior *cluster*; B - Máxima Estrutura Comum entre as moléculas dos 4 primeiros *clusters*.

### Resultados

- Implementação da CAMIS (Clusterização e Análise Molecular com Identificação de Subgrafos), um notebook funcional e executável no Google Colab, compatível com arquivos no formato .sdf;
- Desenvolvimento de uma interface que permite ao usuário ajustar o threshold de similaridade e visualizar os *clusters* formados;
- Identificação do MCS de cada *cluster*, possibilitando a extração de padrões estruturais recorrentes.



Acesse a CAMIS através do QRcode:  
Conheça as funcionalidades da ferramenta.

### Conclusões

O projeto resultou em uma ferramenta prática e acessível para análises em bioinformática, integrando técnicas de clusterização molecular e isomorfismo de grafos em um ambiente interativo. O notebook desenvolvido contribui para otimizar as etapas iniciais do desenvolvimento de fármacos, ao facilitar a identificação de estruturas químicas comuns e a priorização de compostos para testes experimentais. Dessa forma, a solução proposta representa um recurso estratégico para o desenho racional de novos medicamentos, reduzindo custos e acelerando o processo de descoberta farmacêutica.

### Bibliografia

Sydow, D.; Morger, A.; Driller, M.; Volkamer, A. TeachOpenCADD: a teaching platform for computer-aided drug design using open source packages and data. Journal of Cheminformatics, v. 11, art. 29, 2019. DOI: 10.1186/s13321-019-0351-x.