

# Estudo de Inibidores da Anidrase Carbônica $\beta$ da *Mycobacterium tuberculosis* por Modelagem Molecular e Análises de Componentes Principais e de Cluster

## Study of *Mycobacterium tuberculosis* Carbonic Anhydrase $\beta$ Inhibitors by Molecular Modeling and Principal Component and Cluster Analyses

Winnie Mariana A. L. Teixeira,<sup>a</sup> Felipe Leite Coelho da Silva,<sup>b</sup> Carlos Mauricio R. Sant'Anna<sup>\*c</sup>

<sup>a</sup>Universidade Federal Rural do Rio de Janeiro, Instituto de Ciências Exatas, Programa de Pós-Graduação em Modelagem Matemática e Computacional, CEP 23890-000, Seropédica-RJ, Brasil

<sup>b</sup>Universidade Federal Rural do Rio de Janeiro, Instituto de Ciências Exatas, Departamento de Matemática, CEP 23890-000, Seropédica-RJ, Brasil

<sup>c</sup>Universidade Federal Rural do Rio de Janeiro, Instituto de Química, Departamento de Química Fundamental, CEP 23890-000, Seropédica-RJ, Brasil

\*E-mail: [santanna@ufrrj.br](mailto:santanna@ufrrj.br)

Submissão: 12 de Dezembro de 2025 – Aceite: 8 de Abril de 2026 – Publicado online: 23 de Abril de 2026

Tuberculosis is caused by *Mycobacterium tuberculosis*. Although it is a treatable and curable disease, there are cases in which multidrug-resistant strains emerge that are resistant to first- and second-line drugs used in treatment, leading to the need to identify new antibiotic candidates against this mycobacterium. Carbonic anhydrase (CA) is an essential enzyme for the metabolism of many living organisms. In humans, CA is of the  $\alpha$  class, while in bacteria, it is of  $\beta$  class, making it a good candidate molecular target for the development of new antibiotics. In this study, a qualitative model was developed for classifying compounds of the sulfonamide class that showed inhibitory activity on bacterial CA. For this purpose, a comparative model for the 3D structure of *M. tuberculosis* CA was initially constructed, which was used in a molecular docking study with the sulfonamide series. Descriptors resulting from docking and descriptors of isolated molecules were combined with principal component and cluster analysis techniques to generate effective classification models of sulfonamides as inhibitors of *M. tuberculosis* carbonic anhydrase.

**Keywords:** Carbonic anhydrase; *Mycobacterium tuberculosis*; sulfonamides; molecular docking; principal component analysis; cluster analysis.

## 1. Introdução

A tuberculose (TB), causada pelo bacilo *Mycobacterium tuberculosis*, é uma doença infecciosa e transmissível, que afeta principalmente os pulmões, mas que também pode atingir outros órgãos. Segundo o “Global Tuberculosis 2024 report” (relatório global da tuberculose 2024) da Organização Mundial da Saúde (OMS), a TB foi a principal causa de morte por um único agente infeccioso no mundo em 2023. O número global de pessoas diagnosticadas com TB relatado em 2022 foi de 8,2 milhões, com um total de 1,09 milhão de mortes associadas.<sup>1</sup>

No Brasil, segundo o Boletim epidemiológico da TB de março de 2024 do Ministério da Saúde,<sup>2</sup> houve 81.604 casos novos de pessoas diagnosticadas com TB tratadas e mais de 5 mil óbitos. A TB geralmente é curável por meio de tratamento e evitável por meio de medidas preventivas como a imunização pela BCG, tratamento preventivo da tuberculose, que trata a doença em portadores da mico-

bactéria antes de se tornar ativa e pelo controle e prevenção da disseminação.<sup>2,3</sup>

O diagnóstico e o tratamento dessa doença estão disponíveis no SUS (Sistema Único de Saúde) mas, em alguns casos, podem surgir cepas com resistência ao tratamento. Isso ocorre como resultado da constante multiplicação bacteriana em indivíduos que não fazem o tratamento, que fazem o tratamento inadequado ou que interrompem o tratamento antes da cura.<sup>4</sup> A disseminação dos bacilos resistentes torna o tratamento mais custoso, com maior duração e com menor chance de ser eficaz.<sup>2</sup>

Em vista disso, a busca de medicamentos que atuem por novos mecanismos, diferentes daqueles envolvidos na ação dos fármacos já conhecidos, ganha grande importância. A anidrase carbônica (CA) é uma enzima essencial encontrada nos seres vivos. As CAs humana são da classe  $\alpha$ , enquanto as CAs da *M. tuberculosis* (MtCA) são da classe  $\beta$ , que não está presente em mamíferos. As enzimas dos mamíferos são necessárias para o metabolismo intermediário, facilitando

a difusão de  $\text{CO}_2$ , a homeostase do pH e o transporte de íons. A enzima de *M. tuberculosis* está envolvida na sua sobrevivência durante períodos de inanição, hipóxia, condições de estresse no granuloma do hospedeiro, formação de biofilme, virulência e transporte extracelular de DNA; estudos recentes demonstraram inequivocamente que a MtCA é essencial para a sobrevivência de *M. tuberculosis*.<sup>5</sup>

Assim, a inibição da MtCA pode ser uma abordagem promissora para o desenho de agentes antibacterianos com um mecanismo de ação diferente dos agentes farmacológicos clássicos em uso clínico. *M. tuberculosis* contém três genes, Rv1284, Rv3588c e Rv3273, que codificam as isoformas MtCA1, MtCA2 e MtCA3, respectivamente.<sup>6</sup>

A proposta desse trabalho é o uso da modelagem molecular para a geração de descritores para uma série de inibidores conhecidos da MtCA e o estudo para identificar quais desses descritores podem estar relacionados com a atividade inibitória, expressa como constantes de inibição enzimática determinadas experimentalmente.

A modelagem molecular é uma ferramenta comumente usada para auxiliar o desenvolvimento de novos fármacos. Nela, diferentes métodos computacionais são empregados na extração de informações sobre a estrutura molecular, que podem ser usadas em diferentes estratégias, como, por exemplo, o Planejamento de Fármacos Baseado na Estrutura do Alvo (SBDD, do inglês Structure Based Drug Design). Essa estratégia é usada quando se tem o conhecimento da estrutura do alvo, geralmente uma proteína, para o qual se buscam novos ligantes capazes de modular seu funcionamento; por exemplo, se a proteína é uma enzima, podem causar sua inibição.<sup>7</sup>

A otimização da estrutura dos ligantes pode ser feita por métodos que minimizam a energia da estrutura, podendo-se em seguida calcular outras propriedades como volume, densidade eletrônica, energia de orbitais, entre outros.<sup>8</sup> Para se calcular a energia estão disponíveis diferentes modelos, como os métodos clássicos, que incluem a mecânica molecular e a dinâmica molecular, ou os métodos quânticos, entre eles os métodos *ab initio*, os semi-empíricos e os baseados na Teoria do Funcional de Densidade (DFT).<sup>8</sup>

Uma técnica bastante usada no SBDD para definir os modos de interação (as chamadas “poses”) entre um ligante e seu alvo é a docagem molecular, que usa

diferentes algoritmos de busca para se obter as poses, que são em seguida avaliadas através das chamadas funções de pontuação.<sup>7</sup> Essa técnica foi usada nesse estudo.

Para a construção de uma análise qualitativa, a análise estatística foi empregada com os dados gerados pelos métodos da modelagem molecular. Foi usada a análise multivariada de dados, especificamente a análise de componentes principais (ACP) com o objetivo de reduzir a dimensionalidade da matriz de dados para se encontrar combinações lineares que descrevam bem esse conjunto. Os coeficientes dessa combinação linear são os autovetores da matriz covariância, possibilitando um conjunto menor de dados com pouca perda de informação.<sup>9</sup>

Também foi usada a análise de clusters, que agrupa os descritores em conjuntos onde os elementos do grupo possuem características semelhantes e diferentes dos outros conjuntos.<sup>9</sup> É uma ferramenta eficiente para averiguar o conjunto de dados extraídos da ACP a fim de se observar se a separação na clusterização será dada em dois grupos de moléculas, as muito ativas e as pouco ativas, tendo como referência os dados de constante de inibição. Mostrando isso, teremos um conjunto de propriedades em que sua combinação poderá ser usada para se classificar novos ligantes da enzima como candidatos a novos inibidores da MtCA.

## 2. Experimental

O fluxograma representado na Figura 1 mostra as etapas do processo do estudo. Na primeira etapa do estudo foi usada a modelagem molecular para produzir as estruturas moleculares dos inibidores da MtCA e realizar a docagem molecular entre elas; após essa etapa, foram extraídos os descritores para gerar a matriz de dados, discriminados nas Tabelas 4 e 5 desta seção, para se fazer o estudo de ACP e a análise de clusters para identificar os descritores que representem bem o conjunto e mais relacionados com a inibição.

No estudo de docagem molecular foi usada inicialmente a estrutura da isoforma 2 da MtCA (MtCA2) disponível no banco de dados *Protein Data Bank* ([www.rcsb.org/](http://www.rcsb.org/)), com o código 2A5V<sup>10</sup> (disponível em <https://www.rcsb.org/structure/2A5V>), obtida por difração de raios X com

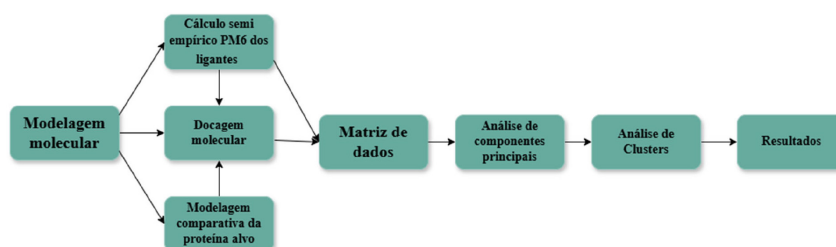


Figura 1. Fluxograma descrevendo as etapas do estudo

resolução de 2,20 Å. No entanto, essa estrutura não se mostrou adequada para o estudo de docagem molecular, por conter um ligante co-cristalizado muito pequeno no seu sítio ativo, o íon tiocianato, o que deixou a cavidade do sítio muito restrita para o recebimento dos inibidores maiores no estudo de docagem. Isso aconteceu porque na docagem os modos de interação dos ligantes (ou poses) são avaliados somente através de mudanças na conformação (variação de ângulos de torção das ligações químicas) e da posição e orientação da molécula no sítio, enquanto a proteína é mantida rígida.

Para se obter uma estrutura da MtCA2 com uma cavidade do sítio mais volumosa, foi usada uma estratégia empregando a modelagem comparativa por meio do servidor Swiss Model (<https://swissmodel.expasy.org/>). Foi usada a sequência primária da estrutura 2A5V<sup>10</sup> e, como molde, a cadeia A da estrutura de código 1G5C<sup>11</sup> no PDB (resolução de 2,10 Å), que é de uma CA de *Methanothermobacter thermautotrophicus* contendo em seu sítio ativo um ligante co-cristalizado mais volumoso, o ácido 4-(2-hidroxietil)-1-piperaziniletanossulfônico (Hepes), uma substância usada para tamponar o meio de cristalização. Embora a identidade global entre as sequências primárias das duas espécies seja baixa (24,2%), a identidade da região que contém os resíduos do sítio ativo (considerando uma distância de até 8,0 Å do íon Zn(II)) é de 30,2%. Portanto, essa região está no limite para obter um modelo aceitável. Dessa forma, obtivemos uma estrutura para a MtCA2 com o sítio ativo em uma conformação mais aberta, que se mostrou adequada para o estudo de docagem. É importante que o sítio ativo, que é onde acontece a interação enzima-ligante, seja suficientemente amplo para que seja possível a interação com ligantes de maior volume. Maiores informações sobre o modelo são apresentadas no material suplementar deste artigo.

Os ligantes explorados nesse estudo foram escolhidos entre inibidores da MtCA com dados de constantes de inibição  $K_i$  relatados na literatura,<sup>6,13</sup> apresentados nas Tabelas 1, 2, 3 e Figura 2.

Entre os compostos há uma série,<sup>13</sup> cuja estrutura geral é apresentada na Figura 2, com os respectivos grupos  $R_1$  e  $R_2$  descritos na Tabela 3.

As estruturas 3D dos compostos foram preparadas com o programa Spartan'24 (Wavefunction, Inc.) e suas energias foram minimizadas usando o método quântico semiempírico PM6 (Parametric Method 6).<sup>14</sup> A análise da interação entre a estrutura da MtCA2 com os compostos foi feita por docagem molecular através do programa GOLD2023.2.0 (CCDC), após adição de átomos de H na estrutura do modelo da MtCA2 conforme procedimento padrão do programa, que considera o pH como neutro. Para o estudo de docagem, as estruturas dos inibidores foram assumidas com o grupo sulfonamida na forma neutra; os inibidores **54**, **55**, **59** e **60** foram usados na forma de ânions após retirada do íon  $\text{Na}^+$

na etapa do desenho molecular. Foram exploradas todas as quatro funções de pontuação disponíveis no programa. Devido à natureza estocástica do algoritmo genético usado pelo GOLD para a busca das poses de interação, fez-se a geração das poses em triplicata, selecionando-se para cada ligante a pose de melhor pontuação da triplicata.

A constante de inibição  $K_i$  quantifica o quanto as moléculas são capazes de inibir a enzima, sendo que quanto menor for  $K_i$ , mais potente é o inibidor. Por sua vez, a pontuação da docagem molecular reflete o quanto a pose de um dado ligante resulta em boas interações com o alvo, ou seja, a enzima. No programa GOLD, cada pose identificada recebe uma pontuação calculada por funções de pontuação, um número adimensional que, quanto mais alto for, melhor será o modo de interação correspondente. O programa GOLD pode quantificar a interação proteína-ligante através de quatro diferentes funções de pontuação: GoldScore, ChemPLP, ChemScore e ASP.<sup>15-18</sup>

Foram analisadas quais propriedades das moléculas isoladas e as de seus complexos com a enzima poderiam estar relacionadas com a atividade inibitória. Foram preparadas matrizes compostas pelos dados de  $K_i$  coletados da literatura e pelos dados calculados nos programas Spartan'24 e GOLD. Essas matrizes contêm 60 objetos, que são os ligantes, e os respectivos conjuntos de variáveis associadas a cada objeto.

As variáveis dependentes são as constantes  $K_i$  e as variáveis independentes são os dados calculados com o programa Spartan'24, que quantificam propriedades físico-químicas dos ligantes, e com o programa GOLD, que quantificam as interações proteína-ligante, incluindo as pontuações e as variáveis que compõem a pontuação de cada função.

A matriz de dados gerada pela docagem molecular passou pela análise multivariada para investigar padrões dos dados e reduzir a dimensionalidade, para se encontrar um conjunto com menos variáveis, mas que separasse bem o conjunto de moléculas mais ativas das menos ativas. A análise de componentes principais (ACP) foi utilizada para redução de dimensionalidade e a análise de *clusters* para verificar quais variáveis dos componentes principais seriam apropriadas para o agrupamento.

A ACP é um método que se aplica a dados multivariados para reduzir a dimensionalidade de modo que as informações importantes da matriz de dados originais sejam mantidas e expressas por um conjunto de variáveis menor. Os componentes principais são combinações lineares desse conjunto onde se tem a maior variabilidade entre os dados, menor perda de informação e sem a multicolinearidade entre os dados.<sup>12,19</sup>

Antes de aplicar o método de ACP, os dados foram padronizados, ou seja, com desvio igual a 1 e média

**Tabela 1.** Conjunto de moléculas e constantes de inibição sobre MtCA2<sup>63</sup>

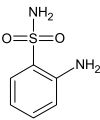
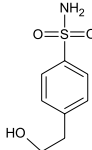
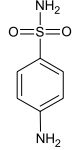
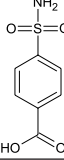
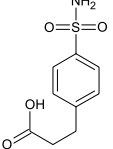
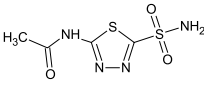
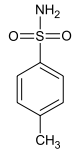
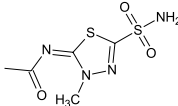
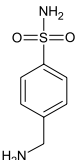
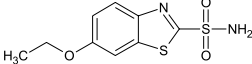
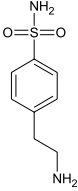
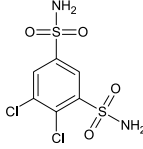
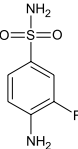
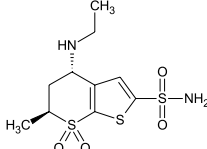
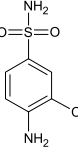
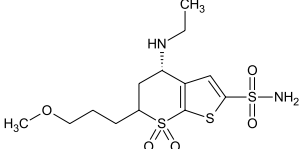
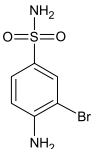
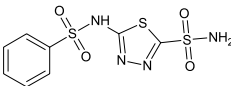
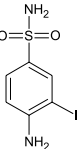
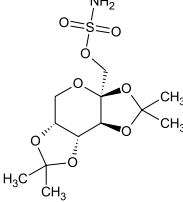
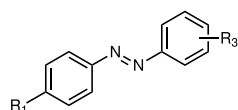
| Composto | Estrutura                                                                           | $K_i$ ( $\mu$ M) | Composto | Estrutura                                                                             | $K_i$ ( $\mu$ M) |
|----------|-------------------------------------------------------------------------------------|------------------|----------|---------------------------------------------------------------------------------------|------------------|
| 1        |    | 33,7             | 21       |    | 34,5             |
| 2        |    | 29,6             | 22       |    | 39,2             |
| 3        |    | 28,4             | 23       |    | 0,009            |
| 4        |    | 38,9             | 24       |    | 0,66             |
| 5        |   | 30,7             | 25       |   | 0,027            |
| 6        |  | 29,1             | 26       |  | 2,01             |
| 7        |  | 28,1             | 27       |  | 0,099            |
| 8        |  | 27,7             | 28       |   | 0,12             |
| 9        |  | 31,6             | 29       |  | 0,467            |
| 10       |  | 32,4             | 30       |  | 0,474            |

**Tabela 1.** Conjunto de moléculas e constantes de inibição sobre MtCA2<sup>6\*</sup> (cont.)

| Composto | Estrutura | $K_i$ ( $\mu$ M) | Composto | Estrutura | $K_i$ ( $\mu$ M) |
|----------|-----------|------------------|----------|-----------|------------------|
| 11       |           | 29,6             | 31       |           | 0,266            |
| 12       |           | 32,5             | 32       |           | 0,717            |
| 13       |           | 2,09             | 33       |           | 0,876            |
| 14       |           | 2,38             | 34       |           | 0,713            |
| 15       |           | 0,978            | 35       |           | 0,682            |
| 16       |           | 3,21             | 36       |           | 0,664            |
| 17       |           | 2,29             | 37       |           | 0,792            |
| 18       |           | 2,63             |          |           |                  |
| 19       |           | 45,2             |          |           |                  |
| 20       |           | 38,3             |          |           |                  |

**Tabela 2.** Conjunto de moléculas e constantes de inibição sobre MtCA2<sup>13</sup>

| Composto | Estrutura | $K_i$ ( $\mu$ M) | Composto | Estrutura | $K_i$ ( $\mu$ M) |
|----------|-----------|------------------|----------|-----------|------------------|
| 38       |           | 0,792            | 44       |           | 0,0211           |
| 39       |           | 0,0044           | 45       |           | 0,0034           |
| 40       |           | 0,0054           | 46       |           | 0,0084           |
| 41       |           | 0,0076           | 47       |           | 0,0046           |
| 42       |           | 0,0042           | 48       |           | 0,0092           |
| 43       |           | 0,0571           | 49       |           | 0,0081           |

**Figura 2.** Série de moléculas extraídas de Abdoli et al., 2023<sup>13</sup>

igual a zero. Assim, as variáveis ficaram na mesma escala para evitar que as variáveis de alta variância afetem o modelo. Os componentes principais (CP) são formados por uma combinação linear onde a primeira componente (Z1) possui maior variância possível e com correlação zero com as demais componentes, e as demais seguem os mesmos critérios.<sup>12,20</sup> Cada CP expressa uma característica do conjunto e o objetivo desse trabalho foi encontrar uma combinação de variáveis que esteja relacionada com a atividade de inibição.

A análise de *cluster* tem o objetivo agrupar as variáveis de acordo com a sua similaridade e separar com maior dissimilaridade possível.<sup>21</sup> O método escolhido foi o particionamento em torno de medoides (PAM, do inglês *Partitioning Around Medoids*) que é mais robusto e menos sensível aos valores atípicos em relação às K-médias

**Tabela 3.** Substituintes dos inibidores da MtCA2 e respectivas constantes de inibição (conforme Figura 2)<sup>13</sup>

| Composto | R <sub>1</sub>                          | R <sub>2</sub>                            | $K_i$ ( $\mu$ M) |
|----------|-----------------------------------------|-------------------------------------------|------------------|
| 50       | OH                                      | <i>p</i> -SO <sub>2</sub> NH <sub>2</sub> | 0,678            |
| 51       | NH <sub>2</sub>                         | <i>p</i> -SO <sub>2</sub> NH <sub>2</sub> | 0,955            |
| 52       | NHMe                                    | <i>p</i> -SO <sub>2</sub> NH <sub>2</sub> | 0,346            |
| 53       | NMe <sub>2</sub>                        | <i>p</i> -SO <sub>2</sub> NH <sub>2</sub> | 5,48             |
| 54       | NHCH <sub>2</sub> SO <sub>3</sub> Na    | <i>p</i> -SO <sub>2</sub> NH <sub>2</sub> | 0,059            |
| 55       | N(Me)CH <sub>2</sub> SO <sub>3</sub> Na | <i>p</i> -SO <sub>2</sub> NH <sub>2</sub> | 0,045            |
| 56       | OH                                      | <i>m</i> -SO <sub>2</sub> NH <sub>2</sub> | 6,48             |
| 57       | NH <sub>2</sub>                         | <i>m</i> -SO <sub>2</sub> NH <sub>2</sub> | 1,98             |
| 58       | NMe <sub>2</sub>                        | <i>m</i> -SO <sub>2</sub> NH <sub>2</sub> | 2,13             |
| 59       | NHCH <sub>2</sub> SO <sub>3</sub> Na    | <i>m</i> -SO <sub>2</sub> NH <sub>2</sub> | 6,55             |
| 60       | N(Me)CH <sub>2</sub> SO <sub>3</sub> Na | <i>m</i> -SO <sub>2</sub> NH <sub>2</sub> | 6,9              |

e mostrou resultados melhores. Esse método busca K objetos representativos (medóides) no conjunto de dados minimizando a soma das distâncias para formar um primeiro conjunto de medoides, que é aprimorado até se chegar ao conjunto ideal para separação.<sup>22</sup>

**Tabela 4.** Descritores moleculares obtidos com o método PM6 (programa Spartan'24)

| Variável                       | Definição                   | Variável                      | Definição                                |
|--------------------------------|-----------------------------|-------------------------------|------------------------------------------|
| Número de tautômeros           |                             | Min. ElPlot (kJ/mol)          | Potencial eletrostático mínimo           |
| Energia (kJ/mol)               |                             | Max. ElPlot (kJ/mol)          | Potencial eletrostático máximo           |
| Energia de solvatação (kJ/mol) |                             | Acc. P-area (Å <sup>2</sup> ) | Área polar acessível ao solvente         |
| HOMO (eV)                      | Energia do HOMO             | Polarizabilidade              |                                          |
| LUMO (eV)                      | Energia do LUMO             | HBD                           | Número de doadores de ligação de H       |
| Momento de dipolo (debye)      |                             | HBA                           | Número de aceptadores de ligação de H    |
| Area (Å <sup>2</sup> )         | Área superficial            | Ovalidade                     | Medida de desvio da forma esférica       |
| Volume (Å <sup>3</sup> )       |                             | Acc. Area (Å <sup>2</sup> )   | Área acessível ao solvente               |
| PSA (Å <sup>2</sup> )          | Área superficial polar      | S° (J/K.mol)                  | Entropia padrão de Formação              |
| H <sub>0</sub> (kJ/mol)        | Entalpia padrão de Formação | G <sub>0</sub> (kJ/mol)       | Energia Livre padrão de Formação         |
| ZPE (au)                       | Energia do ponto zero       | C <sub>v</sub> (J/K)          | Capacidade calorífica a volume constante |

As quantidades de grupos foram determinadas pelo método do cotovelo, que mostrou a quantidade ideal de *clusters* de cada CP, pois mostra a qualidade dos agrupamentos.<sup>23</sup> De antemão foi pensado que seria ideal para o estudo a separação em dois *clusters*, pois permitiria a separação das moléculas em um grupo das muito ativas e outro das pouco ativas; essa técnica confirmou essa hipótese.

### 3. Resultados e Discussão

O modelo da MtCA obtido por modelagem comparativa apresentou boa sobreposição com o molde usado, a CA de *Methanothermobacter thermautotrophicus* contendo o ligante cocristalizado ácido 4-(2-hidroxietil)-1-piperaziniletanossulfônico, com um RMSD baixo (2,31 Å), mostrando que, mesmo com a baixa identidade geral entre as sequências, foi possível obter um modelo com boa semelhança com o molde (Figura 3). Esse modelo, quando sobreposto com a estrutura cristalográfica original da MtCA2 (2A5V), apresentou diferenças conformacionais consideráveis (RMSD = 7,18 Å), principalmente com aumento do volume da cavidade do sítio ativo, como desejado para o estudo de docagem molecular dos inibidores selecionados (Figura 3).

Com esse modelo da MtCA2 foi realizado o estudo de docagem molecular, gerando-se os descritores relacionados às 4 funções de pontuação, que foram reunidos com os descritores das moléculas isoladas. Após se obter os dados de modelagem molecular, os resultados encontrados na ACP, calculados no programa R Studio,<sup>24</sup> passaram por dois critérios. Inicialmente, para se determinar a quantidade

de CP a serem considerados, foi observada a variância. O método do cotovelo mostrou que a partir o sexto componente a proporção da variância explicada pelo CP subsequente cai consideravelmente. Então, foram analisados 5 componentes de cada função; detalhes estatísticos são apresentados nas Tabelas 1S, 2S, 3S e 4S (Informação Suplementar). Incluindo os dados da função GoldScore, foram necessários os primeiros 5 CP para explicar 84% da variância total do conjunto, com a ChemPLP foi preciso 5 CP para explicar 80%, com a ChemScore foi preciso 6 CP para explicar 82,2% e com a função ASP foram necessários 5 CP para expressar 83,9% da variância total.

Os resultados obtidos pela ACP mostraram que os primeiros componentes foram selecionados para a análise de cluster. Com a interpretação dos gráficos de contribuição das variáveis foi possível identificar as variáveis que possuíam os maiores autovetores da matriz correlação (carga ou coeficiente de ponderação). Foram selecionadas as que apresentavam um coeficiente de ponderação maior ou igual a média dos coeficientes.

Os CP geraram novos conjuntos de dados; como cada CP expressa uma característica ou propriedade do conjunto, o propósito foi encontrar um conjunto de dados que esteja associado à inibição da MtCA, que é expressa pelos dados de  $K_i$ . Cada CP foi submetido à análise de clusters (PAM) com a finalidade de se observar o agrupamento e verificar se algum conjunto com a dimensionalidade reduzida conseguiria a relação entre os subgrupos e  $K_i$ , a variável dependente.

Pode-se observar nos gráficos dos clusters (Figuras 4 a 7) que a partir das variáveis dos primeiros componentes de cada

**Tabela 5.** Descritores obtidos com o programa GOLD 2023.2.0

| Variáveis             | Definição                                                                | Variáveis                            | Definição                                                                                                      |
|-----------------------|--------------------------------------------------------------------------|--------------------------------------|----------------------------------------------------------------------------------------------------------------|
| GoldScore             |                                                                          | ChemPLP                              |                                                                                                                |
| <b>GoldScore</b>      | Maior pontuação da triplicata das docagens feitas com a função           | <b>PLP.Fitness</b>                   | Maior pontuação da triplicata das docagens                                                                     |
| <b>Hbond.ext</b>      | Energia de ligação de hidrogênio proteína-ligante                        | <b>ChemScore.HBond</b>               | Contribuição de ligação de hidrogênio proteína-ligante (ChemScore)                                             |
| <b>VDW.Ext</b>        | Energia de Van der Waals proteína-ligante                                | <b>ChemScore.Metal</b>               | Contribuição de ligação com metais (ChemScore)                                                                 |
| <b>torsion.int</b>    | Energia de deformação torcional do ligante                               | <b>PLP.ligand.clash</b>              | Contribuição de colisão proteína-ligante (PLP)                                                                 |
| <b>VDW.int</b>        | Energia de van der Waals do ligante                                      | <b>PLP.ligand.torsion</b>            | Penalidade de tensão torsional interna do ligante (PLP)                                                        |
| <b>DG</b>             | Energia livre                                                            | <b>PLP.Part.buried</b>               | Contribuição de interações oclusas (PLP)                                                                       |
| <b>HBond</b>          | Contribuição de ligação de hidrogênio proteína-ligante                   | <b>PLP.Part.Hbond</b>                | Contribuição de ligações de hidrogênio (PLP)                                                                   |
| <b>Lipo</b>           | Contribuição lipofílica proteína-ligante                                 | <b>PLP.Part.metal</b>                | Contribuição de ligação com metais (PLP)                                                                       |
| <b>Metal</b>          | Contribuição de ligação com metais                                       | <b>PLP.Part.nonpolar</b>             | Contribuição de interação apolar (PLP)                                                                         |
|                       |                                                                          | <b>PLP.Part.repulsive</b>            | Contribuição de interação repulsiva (PLP)                                                                      |
|                       |                                                                          | <b>PLP.Protein.Rotated.Torsion</b>   | Contribuição de torção de ângulos da proteína (PLP)                                                            |
|                       |                                                                          | <b>ChemScore.CHOScore</b>            | Contribuição de ligações de hidrogênio fracas CH---O                                                           |
|                       |                                                                          | <b>ChemScore.Internal.Correction</b> | Contribuição de colisão interna e de torção do ligante                                                         |
| Chemscore             |                                                                          | ASP                                  |                                                                                                                |
| <b>Chemscore</b>      | Maior pontuação da triplicata das docagens feitas com a função Chemscore | <b>ASP Fitness</b>                   | Maior pontuação da triplicata das docagens feitas com a função ASP                                             |
| <b>DEClash</b>        | Penalidade de colisão proteína-ligante                                   | <b>ASP</b>                           | Potencial estatístico calculado mais o termo de colisão ChemScore e o termo de energia interna                 |
| <b>DEInternal</b>     | Penalidade de deformação de torção do ligante                            | <b>DEClash</b>                       | Penalidade de colisão proteína-ligante para o valor ASP                                                        |
| <b>DG</b>             | Variação de energia livre                                                | <b>DEInternal</b>                    | Contribuição de ligação de hidrogênio intramolecular do ligante para o valor ASP                               |
| <b>HBond</b>          | ligação de hidrogênio proteína-ligante                                   | <b>internal.correction</b>           | Deslocamento da energia interna do ligante                                                                     |
| <b>Lipo</b>           | Contribuição lipofílica proteína-ligante                                 | <b>ASP.map</b>                       | Potencial estatístico total calculado como uma soma de todas as combinações de átomos da proteína e do ligante |
| <b>Metal</b>          | Ligação de metais                                                        | <b>RotatedTorsions</b>               | Contribuição Termo de congelamento de ligação rotativa para o valor ASP                                        |
| <b>Part.repulsive</b> | Interação de repulsão                                                    |                                      |                                                                                                                |

função as moléculas puderam ser separadas em 2 grupos, um de moléculas com  $K_i \leq 3 \mu\text{M}$  e outro de moléculas com  $K_i \geq 3 \mu\text{M}$ . Pode-se observar que os grupos não se sobrepõem, mostrando que houve boa separação. Também pode ser visto nos eixos as dimensões que representam a variância dos elementos e assim vemos que dentro dos grupos há uma proximidade dos elementos, evidenciando a similaridade entre eles, como se pode observar no eixo Y. Os gráficos dos clusters mostram também as moléculas que tiveram uma distância maior, evidenciando aquelas com características mais diferentes nesse cluster, como pode se ver pelo eixo x.

Combinando as duas técnicas, foi possível identificar que existem subconjuntos dos dados originais, ou seja, com menor dimensionalidade, que caracterizam bem o grau de inibição da enzima, evidenciados pelos clusters. Através da docagem obteve-se o nível de interação inibidor-enzima por meio das pontuações de cada função. Pela combinação dos descritores das funções de pontuação da docagem com os descritores das moléculas isoladas, observou-se que com quando os descritores das funções GoldScore, ChemPLP e ASP foram usados, houve mais de 85% de acertos na separação dos compostos nos critérios qualitativos de

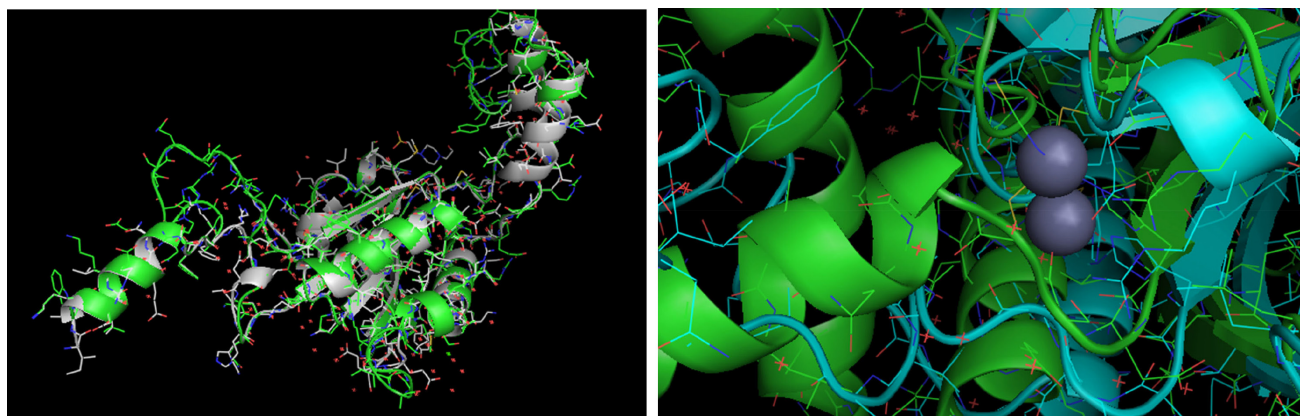


**Figura 3.** Sobreposição do modelo comparativo da CA de *M. tuberculosis* (verde) com o molde, a estrutura da CA de *Methanothermobacter thermautotrophicus* (código no PDB: 1G5C, em branco). Sobreposição do sítio catalítico da estrutura cristalográfica de MtCA (código no PDB: 2A5V, em verde) com o modelo da MtCA (em ciano) gerado por modelagem comparativa a partir do molde 1G5C; íons Zn(II) localizados no sítio catalítico representados como esferas de cor cinza (direita). Figuras geradas com o programa PyMol v 0.99 (DeLano Scientific LLC)

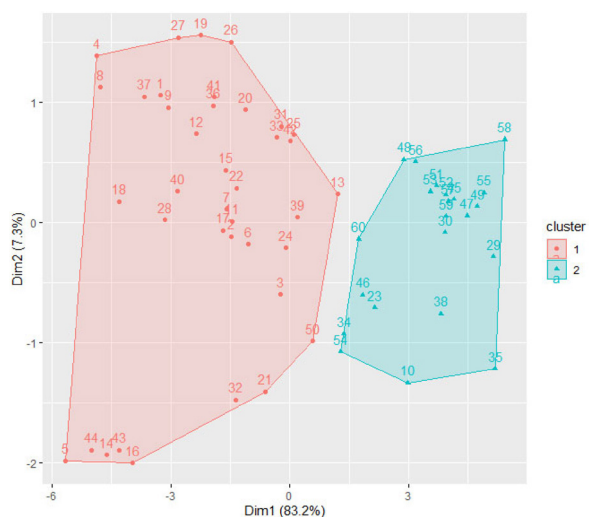


**Figura 4.** Gráfico de clusters do CP1 da função GoldScore (os números representam cada molécula numerada de 1 a 60)

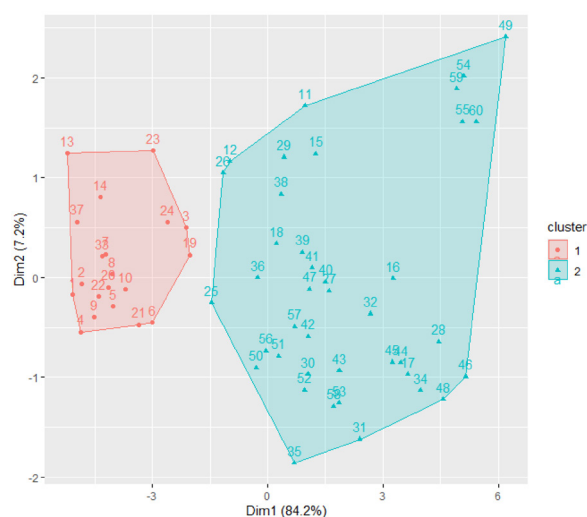


**Figura 6.** Gráfico de clusters do CP1 da função ChemScore (os números representam cada molécula numerada conforme as Tabelas 1, 2 e 3)

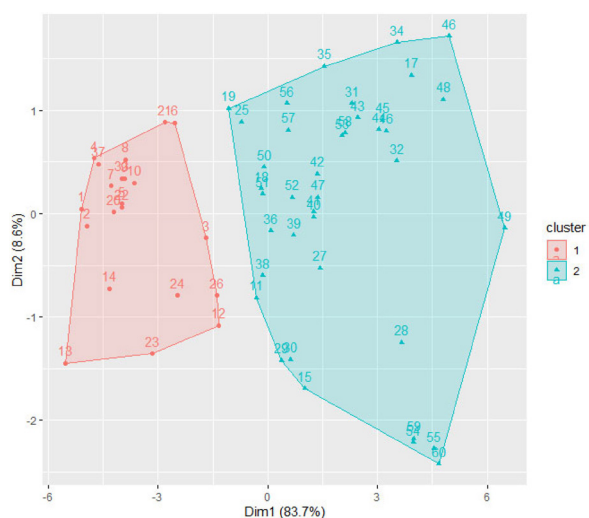


**Figura 5.** Gráfico de clusters do CP1 da função ChemPLP (os números representam cada molécula numerada conforme as Tabelas 1, 2 e 3)

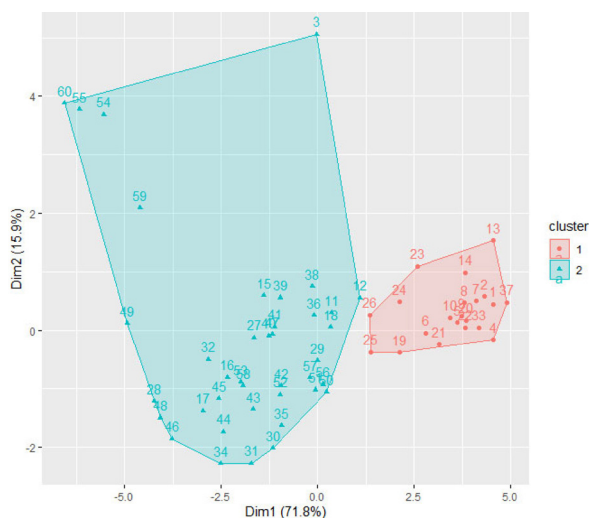


**Figura 7.** Gráfico de clusters do CP1 da função ASP (os números representam cada molécula numerada conforme as Tabelas 1, 2 e 3)

atividade de inibição da enzima, mostrando que esses métodos apresentam um ótimo potencial classificatório para o sistema em estudo (Tabela 6). O uso da função ChemScore não apresentou uma proporção de acerto satisfatória, indicando que para esse conjunto de dados e com as técnicas usadas essa função não apresentou soluções tão boas quanto as outras.

**Tabela 6.** Porcentagem de acertos na separação dos clusters nos componentes principais

| CP  | GoldScore | ChemPLP | ChemScore | ASP   |
|-----|-----------|---------|-----------|-------|
| CP1 | 86,6%     | 85%     | 40%       | 86,6% |
| CP2 | 31,6%     | 33,3%   | 31,6%     | 31,6% |
| CP3 | 58,3%     | 53,3%   | 40%       | 68,3% |
| CP4 | 51,6%     | 51,6%   | 46,6%     | 50%   |
| CP5 | 48,3%     | 56,6%   | 50%       | 31,6% |

O primeiro CP da função GoldScore foi expresso por:

$$CP_1 = 0,265 \cdot x_1 + 0,263 \cdot x_2 + 0,261 \cdot x_3 + 0,257 \cdot x_4 + 0,256 \cdot x_5 + 0,286 \cdot x_6 + 0,25 \cdot x_7 + 0,238 \cdot x_8 + 0,243 \cdot x_9 + 0,253 \cdot x_{10} + 0,218 \cdot x_{11} + 0,215 \cdot x_{12} + 0,194 \cdot x_{13}$$

onde  $x_1$  é a ovalidade;  $x_2$ , área superficial;  $x_3$ , polarizabilidade;  $x_4$ ,  $S_0$ ;  $x_5$ , volume;  $x_6$ ,  $C_v$ ;  $x_7$ , área acessível ao solvente;  $x_8$ , VDW.Ext;  $x_9$ , pontuação GoldScore;  $x_{10}$ , VDW ponderada;  $x_{11}$ , área polar;  $x_{12}$ , HBA e  $x_{13}$ , ZPE.

Para a função ChemPLP:

$$CP_1 = 0,265 \cdot x_1 + 0,263 \cdot x_2 + 0,263 \cdot x_3 + 0,262 \cdot x_4 + 0,261 \cdot x_5 + 0,26 \cdot x_6 + 0,259 \cdot x_7 - 0,24 \cdot x_8 - 0,231 \cdot x_9 + 0,219 \cdot x_{10} + 0,213 \cdot x_{11} + 0,206 \cdot x_{12} + 0,197 \cdot x_{13}$$

onde  $x_1$  é a área superficial;  $x_2$ , ovalidade;  $x_3$ , polarizabilidade;  $x_4$ , volume;  $x_5$ ,  $S_0$ ;  $x_6$ ,  $C_v$ ;  $x_7$ , área acessível ao solvente;  $x_8$ , área polar acessível ao solvente;  $x_9$ , PLP.PLP;  $x_{10}$ , PLP.Fitness;  $x_{11}$ , área polar;  $x_{12}$ , HBA e  $x_{13}$ , ZPE.

E, para a função função ASP:

$$CP_1 = 0,261 \cdot x_1 + 0,26 \cdot x_2 + 0,259 \cdot x_3 + 0,257 \cdot x_4 + 0,261 \cdot x_5 + 0,26 \cdot x_6 + 0,254 \cdot x_7 - 0,254 \cdot x_8 + 0,252 \cdot x_9 + 0,25 \cdot x_{10} + 0,221 \cdot x_{11} + 0,218 \cdot x_{12} + 0,188 \cdot x_{13}$$

onde  $x_1$  é a ovalidade;  $x_2$ ,  $S_0$ ;  $x_3$ , área superficial;  $x_4$ ,  $C_v$ ;  $x_5$ , polarizabilidade;  $x_6$ , volume;  $x_7$ , ASP.ASP;  $x_8$ , ASP.map;  $x_9$ , pontuação ASP;  $x_{10}$ , área acessível ao solvente;  $x_{11}$ , área polar;  $x_{12}$ , HBA e  $x_{13}$ , ZPE.

Assim, do conjunto inicial, foi possível identificar as variáveis mais relevantes na diferenciação da atividade dos compostos. O resultado mostrou que a combinação

das pontuações de diferentes funções de docagem com propriedades moleculares se mostraram frequentemente relevantes na classificação da atividade dos compostos. Em cada modelo CP com boa capacidade de classificação as pontuações respectivas (e alguns de seus componentes) foram incluídas como termos para as funções GoldScore, ChemPLP e ASP. Dentre os termos que descrevem propriedades moleculares, aparecem com frequência a ovalidade, área, volume, polarizabilidade e capacidade de receber ligações de hidrogênio.

De fato, uma análise geral das estruturas indica que o aumento do volume molecular (e, conseqüentemente da área superficial) parece ser benéfico para o aumento da atividade, assim como grupos capazes de receber ligações de hidrogênio, como pode ser visto nos valores de  $K_i$  dos compostos da tabela 2, que estão entre os menores do conjunto de compostos, boa parte na faixa nM. Esses compostos contêm o grupo carbamotioil-amida em sua parte central, capaz de atuar como receptor de ligações de hidrogênio, e o aumento da cadeia hidrofóbica parece ser, em geral, benéfico para a atividade da série. A presença da ovalidade entre os termos indica que a forma, e não somente o volume, das moléculas influencia a atividade; novamente, vemos que as estruturas da tabela 2, as mais potentes, apresentam um perfil mais alongado, em comparação com as demais, com uma extremidade de caráter hidrofóbico no lado oposto ao do grupo sulfonamida.

## 4. Conclusão

Neste trabalho foram combinados métodos de modelagem molecular, ACP e análise de clusters para encontrar um conjunto de dados de baixa dimensionalidade que esteja associado à classificação qualitativa da atividade inibitória sobre a CA  $\beta$  da *M. tuberculosis* de 60 sulfonamidas selecionadas da literatura.

Além dos dados experimentais de atividade das sulfonamidas, a matriz de dados foi preparada tanto com descritores obtidos por modelagem molecular como variáveis relacionadas às moléculas, obtidos por cálculos quânticos semiempíricos, quanto descritores relacionados à interação receptor-ligante, obtidos por docagem molecular, que simulou a interação entre os inibidores e um modelo 3D da enzima gerado por modelagem comparativa. A quantidade de descritores encontrados estava entre 34 e 40 dependendo da função de pontuação de docagem escolhida, o que resultou em uma matriz de dados com grande quantidade de variáveis. A técnica de componentes principais se mostrou eficaz na redução da dimensionalidade pois encontrou conjuntos de variáveis em menor quantidade sem perda de informação. Em seguida, a análise de clusters mostrou quais componentes principais geravam subgrupos de descritores que estavam bem relacionados com a

classificação da inibição em muito ativas ou pouco ativas, chegando a uma separação com 85% ou mais de acerto quando os descritores moleculares eram combinados com os das funções de pontuação GoldScore, ChemPLP e ASP.

Os resultados deste estudo confirmaram que a modelagem molecular combinada com as técnicas de componentes principais e de análise de clusters podem ser úteis para classificação de moléculas como inibidores da enzima estudada. Isso sugere, portanto, que o procedimento pode ser explorado em estudos de triagem virtual de compostos do grupo das sulfonamidas para se encontrar candidatos a inibidores da CA  $\beta$  da *M. tuberculosis*, com potencial para o desenvolvimento de novos fármacos para o tratamento da tuberculose.

## Informações Suplementares

As tabelas de dados estão disponíveis gratuitamente em <https://rvq.s bq.org.br/pdf/AR2026-5126-MS>.

## Agradecimentos

Os autores agradecem à CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico (C.M.R.S., Processo 309354/2025-0; F.L.C.S., Processo 312646/2025-9) e à FAPERJ – Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro – pelo apoio fornecido ao longo deste trabalho.

## Referências Bibliográficas

1. Site World Health Organization. *Global Tuberculosis Report 2024*. Disponível em: <<https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024>>. Acesso em: 14 abril 2026.
2. Site Ministério da Saúde. *Boletim Epidemiológico: Tuberculose 2024*. Disponível: <<https://www.gov.br/aids/pt-br/central-de-conteudo/boletins-epidemiologicos/2024/boletim-epidemiologico-tuberculose-2024>>. Acesso em: 14 abril 2026.
3. Brasil. Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância das Doenças Transmissíveis. *Brasil livre da tuberculose: Plano nacional pelo fim da tuberculose como problema de saúde pública*. Editora MS: Brasília, 2017.
4. Brasil. Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis. *Brasil livre da tuberculose: Plano nacional pelo fim da tuberculose como problema de saúde pública – Estratégias para 2021–2025*. Editora MS: Brasília, 2021.
5. Degiacomi, G.; Gianibbi, B.; Recchia, D.; Stelitano, G.; Truglio, G. I.; Marra, P.; Stamilla, A.; Makarov, V.; Chiarelli, L. R.; Manetti, F.; Pasca, M. R.; CanB, a Druggable Cellular Target in *Mycobacterium tuberculosis*. *ACS Omega* **2023**, *8*, 25209. [Crossref]
6. Carta, F.; Maresca, A.; Covarrubias, A. S.; Mowbray, S. L.; Jones, T. A.; Supuran, C. T.; Carbonic Anhydrase Inhibitors: Characterization and Inhibition Studies of the Most Active  $\beta$ -Carbonic Anhydrase from *Mycobacterium Tuberculosis*, rv3588c. *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 6649. [Crossref]
7. Barreiro, E. J.; Fraga, C. A. M.; *Química Medicinal: As Bases Moleculares da Ação dos Fármacos*, 3a. ed., Artmed Editora: Porto Alegre, 2015.
8. Sant'Anna, C. M. R.; Métodos de Modelagem Molecular para Estudo e Planejamento de Compostos Bioativos: Uma Introdução. *Revista Virtual de Química* **2009**, *1*, 49. [Crossref]
9. Lattin, J.; Carroll, J. D.; Green, P. E. *Análise de Dados Multivariados*, 1a. ed., Cengage Learning: São Paulo, 2011.
10. Covarrubias, A. S.; Bergfors, T.; Jones, T. A.; Hogbom, M.; Structural Mechanics of the pH-dependent Activity of  $\beta$ -Carbonic Anhydrase from *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* **2006**, *281*, 4993. [Crossref]
11. Strop, P.; Smith, K.; Iverson, T; Crystal Structure of the “cab”-type  $\beta$  Class Carbonic Anhydrase from the Archaeon *Methanobacterium thermoautotrophicum*. *Journal of Biological Chemistry* **2001**, *276*, 10299. [Crossref]
12. Izbicki, R.; Santos, T. M.; *Aprendizado de Máquina: Uma Abordagem Estatística*, 1a. ed., Rafael Izbicki: São Paulo, 2020.
13. Abdoli, M.; Bonardi, A.; Paoletti N.; Aspatwar, A.; Parkkila, S.; Gratteri, P.; Supuran, C. T.; Žalubovskis, R.; Inhibition studies on human and mycobacterial carbonic anhydrases with N-((4-sulfamoylphenyl) carbamothioyl) amides. *Molecules* **2023**, *28*, 4020. [Crossref]
14. Stewart, J. J.; Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **2007**, *13*, 1173. [Crossref]
15. Kirton, S. B.; Murray, C. W.; Verdonk, M. L.; Taylor, R. D.; Prediction of Binding Modes for Ligands in the Cytochromes P450 and Other Heme-Containing Proteins. *Proteins: Structure, Function, and Bioinformatics* **2008**, *58*, 836. [Crossref]
16. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P.; Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *Journal of Computer-Aided Molecular Design* **1997**, *11*, 425. [Crossref]
17. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D.; Flexible Docking Using Tabu Search and an Empirical Estimate of Binding Affinity. *Proteins: Structure, Function, and Genetics* **1998**, *33*, 367. [PubMed]
18. Mooij, W. T.; Verdonk, M. L.; General and Targeted Statistical Potentials for Protein–Ligand Interactions. *Proteins* **2005**, *61*, 272. [Crossref]

19. Hongyu, K.; Sandanielo, V. L. M.; de Oliveira Júnior, G. J.; Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação. *E&S Engineering and Science* **2016**, *5*, 83. [[Crossref](#)]
20. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; *An Introduction to Statistical Learning*, Springer: New York, 2013.
21. Kaufman, L.; Rousseeuw, P. J.; *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons: New Jersey, 2009.
22. Wehrens, R.; *Chemometrics with R: Multivariate Data Analysis in the Natural and Life Sciences*, 3a. ed., Springer: New York, 2011.
23. Bholowalia, P.; Kumar, A.; EBK-Means: A Clustering Technique Based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications* **2014**, *105*, 17. [[Link](#)]
24. Posit Software. Disponível em: <<https://posit.co/>>. Acesso em: 14 abril 2026.