

**UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE  
PORTO ALEGRE – UFCSPA  
CURSO DE PÓS-GRADUAÇÃO EM BIOCÊNCIAS**

**Aureo André Karolczak**

**Desenvolvimento de Modelos QSAR  
para Identificar Inibidores da Enzima  
Enoil-ACP-Redutase de  
*Mycobacterium tuberculosis***

**UFCSPA**  
Universidade Federal de Ciências da Saúde  
de Porto Alegre

**Porto Alegre  
2022**

**Aureo André Karolczak**

**Desenvolvimento de Modelos QSAR  
para Identificar Inibidores da Enzima  
Enoil-ACP-Redutase de  
*Mycobacterium tuberculosis***

Dissertação submetida ao Programa de Pós-Graduação em Biociências da Fundação Universidade Federal de Ciências da Saúde de Porto Alegre como requisito para a obtenção do grau de Mestre.

Orientador: Prof. Dr. Rafael Andrade Caceres

Coorientador: Prof. Dr. Luis Fernando S. M. Timmers

**Porto Alegre  
2022**

#### Catálogo na Publicação

Karolczak, Aureo André  
Desenvolvimento de Modelos QSAR para Identificar  
Inibidores da Enzima Enoil-ACP-Redutase de Mycobacterium  
tuberculosis / Aureo André Karolczak. -- 2022.  
70 p. : il., graf., tab. ; 30 cm.

Dissertação (mestrado) -- Universidade Federal de  
Ciências da Saúde de Porto Alegre, Programa de  
Pós-Graduação em BioCiências, 2022.

Orientador(a): Rafael Andrade Caceres ;  
coorientador(a): Luis Fernando S. M. Timmers.

1. Modelagem molecular. 2. Tuberculose. 3. QSAR. 4.  
CoMFA. 5. InhA. I. Título.

Sistema de Geração de Ficha Catalográfica da UFCSPA com os dados  
fornecidos pelo(a) autor(a).

## **INSTITUIÇÕES E FONTES FINANCIADORAS**

Universidade Federal de Ciências da Saúde de Porto Alegre – UFCSPA  
Programa de Pós Graduação em Biociências - UFCSPA

## **AGRADECIMENTOS**

Agradeço ao meu orientador, Prof. Rafael Andrade Caceres, por apoiar essa pesquisa e por sempre me incentivar, mesmo nos momentos mais complicados da pandemia do COVID-19. Obrigado, Professor!

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	11
1.1 A TUBERCULOSE .....	11
1.2 O BACILO <i>Mycobacterium tuberculosis</i> , .....	14
<b>1.2.1 Origem e evolução</b> .....	15
<b>1.2.2 Patologia</b> .....	17
1.3 ALVO FARMACOLÓGICO E AÇÃO NO SÍTIO CATALÍTICO .....	18
1.4 AÇÃO DE MEDICAMENTOS E MECANISMOS DE RESISTÊNCIA NO SÍTIO DE INTERESSE.....	25
1.5 O MODELO QSAR.....	28
<b>1.5.1 QSAR clássico, o modelo Hansh-Fujita</b> .....	30
<b>1.5.2 QSAR 3D CoMFA</b> .....	30
<b>2 OBJETIVOS</b> .....	33
2.1 OBJETIVO GERAL .....	33
2.2 OBJETIVOS ESPECÍFICOS .....	33
<b>3 METODOLOGIA</b> .....	34
3.1 ESTUDO SAR.....	34
3.2 FUNÇÃO QSAR CLÁSSICO .....	37
3.3 CoMFA.....	40
<b>3.3.1 Alinhamento</b> .....	40
<b>3.3.2 Geração do modelo</b> .....	40
3.5 DOCAGEM MOLECULAR .....	41
<b>4 RESULTADOS</b> .....	43
<b>5 DISCUSSÕES</b> .....	43
<b>6 CONCLUSÕES</b> .....	62
<b>REFERÊNCIAS</b> .....	63
<b>ANEXO I – APROVAÇÃO DO COMITÊ DE PESQUISA</b> .....	68
<b>CURRICULO LATES</b> .....	69

## LISTA DE ABREVIATURAS E SIGLAS

AG – Ácidos Graxos

AM – Ácidos micólicos

CoMFA – *Comparative Molecular Field Analysis*

DM – Descritores Moleculares

EBEF – *Estimated Free Energy of Binding*

ETH – Etionamida

FAT - Fatty Acid Synthase

INH – Isoniazida

InhA – enoil-ACP-redutase

Log P – Coeficiente de partição água-octanol

MLR – *Multiple Linear Regression*

MTB – *Mycobacterium tuberculosis*

NCEs – Novas entidades químicas

OMS – Organização Mundial da Saúde

P&D – Pesquisa e Desenvolvimento

PLS – Mínimos quadrados parciais

QSAR – *Quantitative Structure-Activity Relationships*

QSAR-3D – *Quantitative Structure-Activity Relationships Tridimensional*

RMS – *Root Mean Square*

SAR – *Relação estrutura atividade*

TB – Tuberculose

TB-MDR – Tuberculose multidroga resistente

TCL – Triclosan

## LISTA DE ILUSTRAÇÕES

Gráfico 1 - Relação entre incidência da Tuberculose e renda.....	11
Gráfico 2 - Causas de mortes em países de alta renda.....	12
Gráfico 3 - Causas de morte em países de baixa renda.....	12
Tabela 1 - Novos produtos terapêuticos aprovados ou recomendados por categoria de doença (2000 – 2011).....	13
Tabela 2 - Novos Produtos Terapêuticos para Tratamento da Tuberculose (2000-2011) .....	14
Figura 1 - Migração hipotética de linhagens ancestrais (preto, verde, marrom e lilás) e modernas (vermelho, azul e roxo) .....	15
Figura 2 - Árvore filogenética do <i>Mycobacterium tuberculosis</i> . .....	16
Figura 3 - Estrutura e constituintes do granuloma de <i>M. tuberculosis</i> .....	18
Figura 4 - A enzima InhA, evidenciando em detalhe o cofator NDAH (amarelo) no sítio catalítico.....	19
Figura 5 - Ciclo FAS II. No destaque em vermelho, a enzima InhA.....	20
Figura 6 - Composição da parede celular da MTB .....	21
Figura 7 - Ácido micólico.....	21
Figura 8 - Interações proteína-proteína conhecidas e previstas entre a InhA e outras enzimas.....	22
Figura 9 - Pocket do sitio catalítico da InhA.....	23
Figura 10 - Mecanismo proposto para redução catalisada pela enoil-ACP-redutase InhA do substrato enoil com o cofator NADH.....	23
Figura 11 - Interações moleculares entre a molécula CID 447767 e a enzima InhA. Depositado sob código PDB 1P44.....	24
Figura 12 - Ativação do pró-fármaco INH via enzima KatG .....	25
Figura 13 -Ligação da INH ativada ao cofator NAD <sup>+</sup> formando o aduto INH-NAD. ...	26
Figura 14 - Ativação do pró-fármaco ETH pela enzima EthA. ....	26
Figura 15 - Ligação da ETH ativada ao cofator NAD <sup>+</sup> formando o aduto ETH-NAD. ...	26
Figura 16 - Ligação de hidrogênio entre TCL e TYR 158 .....	27
Tabela 3 - Mutações de interesse em InhA. ....	28
Figura 17 - GRID 3D para cálculo de energias eletrostáticas e estéricas. ....	32

Figura 18 - Fluxograma para construção dos modelos matemáticos QSAR clássico e CoMFA.....	34
Tabela 4 - Resultados do estudo de SAR promovido por He; Alian; Montellano. ....	35
Figura 19 - Núcleo comum de série de análogos de arilamidas do estudo de SAR...	36
Figura 20 - Divisão do data set em training set e test set.....	37
Tabela 5 - O conjunto de moléculas que compõem o estudo de SAR, divididas em test set e training set, e identificadas pelo código CID utilizado na base dados PubChem. ....	39

## RESUMO

A tuberculose é motivo de preocupação mundial devido a sua prevalência em países em desenvolvimento, e pela capacidade da micobactéria de apresentar resistência aos esquemas terapêuticos existentes. Nesse contexto, propôs-se a modelagem QSAR (*Quantitative Structure-Activity Relationships*), como forma de avaliar, *in silico*, a atividade inibitória de moléculas candidatas às etapas de aperfeiçoamento molecular e/ou ensaios *in vitro*, reduzindo tempo de pesquisa e custos financeiros. Para tanto, foi utilizado o estudo de SAR (*Structure-Activity Relationships*) conduzido por He, Alian e Montellano (2007) sobre uma série de arilamidas, testadas como inibidoras da enzima enoil-ACP-redutase (InhA) de *Mycobacterium tuberculosis*. Foram desenvolvidos os modelos QSAR Hansh-Fujita (clássico) e CoMFA (*Comparative Molecular Field Analysis*). O modelo QSAR clássico obteve o melhor resultado estatístico, utilizando *Multiple Linear Regression* (MLR), sendo a validação interna com fator de correlação  $R^2$  de 0,9012 e qualidade preditiva, conforme o indicador de Stone-Geisser  $Q^2$  de 0,8612. Já na validação externa foi obtido um fator de correlação  $R^2$  de 0,9298 e um  $Q^2$  de 0,720, indicando um modelo matemático altamente preditivo. O Modelo CoMFA obteve um  $Q^2=0,6520$  na validação interna, o que permitiu estimar os campos energéticos no entorno das moléculas utilizadas, informação essencial para propor aperfeiçoamento molecular. Construiu-se uma biblioteca de pequenas moléculas, com análogos aos utilizados no estudo de SAR, que foram submetidos à função QSAR clássico, tendo com resultado um grupo de dez moléculas com elevada atividade biológica estimada. Os resultados da docagem molecular sugerem que os dez análogos apontados pelo modelo QSAR clássico apresentaram energia livre de ligação estimada favorável. A conclusão aponta a metodologia QSAR como uma ferramenta eficiente e eficaz na busca de moléculas *drug-like* promissoras.

**Palavras-chave:** Tuberculose. InhA. QSAR. CoMFA. Docagem.

## ABSTRACT

Tuberculosis is a matter of global concern due to its prevalence in developing countries and the ability of mycobacteria to present resistance to existing therapeutic regimens. In this context, QSAR (Quantitative Structure-Activity Relationships) modeling was proposed as a way of evaluating, *in silico*, the inhibitory activity of candidate molecules for molecular improvement and/or *in vitro* assays, reducing research time and financial costs. For this purpose, the SAR (Structure-Activity Relationships) study conducted by He, Alian and Montellano (2007) was used on a series of arylamides, tested as inhibitors of the enzyme enoyl-ACP-reductase (InhA) of *Mycobacterium tuberculosis*. The Hansh-Fujita (classical) and CoMFA (Comparative Molecular Field Analysis) QSAR models were developed. The classic QSAR model obtained the best statistical result, using Multiple Linear Regression (MLR), with internal validation with correlation factor  $R^2 = 0.9012$  and predictive quality, according to the Stone-Geisser indicator  $Q^2 = 0.8612$ . In the external validation, a correlation factor  $R^2 = 0.9298$  and a  $Q^2 = 0.720$  was obtained, indicating a highly predictive mathematical model. The CoMFA Model obtained a  $Q^2 = 0.6520$  in the internal validation, which allowed estimating the energy fields around the molecules used, essential information to propose molecular improvement. A library of small molecules was built, with analogs to those used in the SAR study, which were subjected to classic QSAR function, resulting in a group of ten molecules with high estimated biological activity. The docking results suggest that the ten analogs pointed out by the classical QSAR model presented low estimated binding free energy. The conclusion points to the QSAR methodology as an efficient and effective tool in the search for promising drug-like molecules.

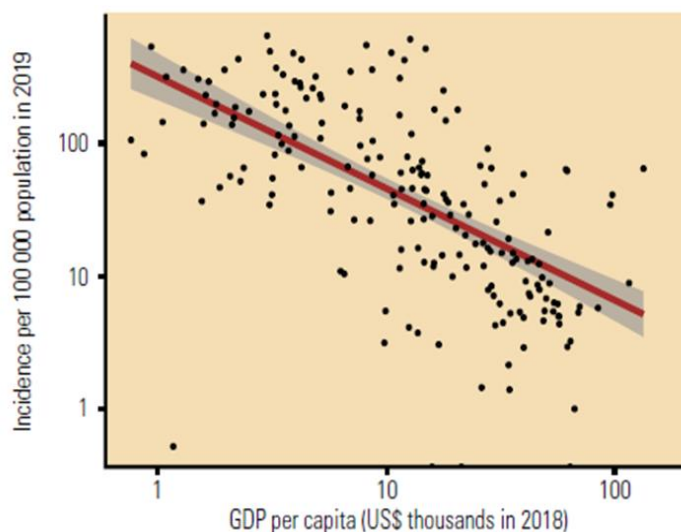
**Keywords:** Tuberculosis. InhA. QSAR. CoMFA. Docking

## 1 INTRODUÇÃO

### 1.1 A TUBERCULOSE

O *Global Tuberculosis Report 2020* coloca a tuberculose (TB) como uma das dez principais causas de morte em todo o mundo, e a principal causa de morte provocada por um único agente de infecção. Outro dado relevante é que aproximadamente um quarto da população mundial está infectado com o bacilo *Mycobacterium tuberculosis* (MTB). O estudo conduzido pela Organização Mundial da Saúde (OMS) aponta uma correlação entre a pobreza e a incidência da enfermidade. O gráfico 1 sugere que quanto maior a renda *per capita* menor será a incidência da tuberculose, deixando claro o grave aspecto social que envolve essa doença.

Gráfico 1 - Relação entre incidência da Tuberculose e renda

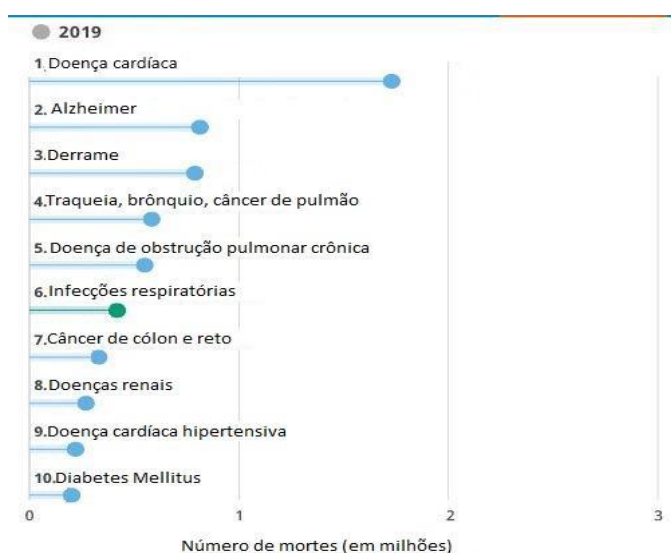


Fonte: *Global Tuberculosis Report 2020*.

A TB é considerada parte das chamadas Doenças Negligenciadas (Pedrique *et al.* 2013) e que se relacionam com a pobreza, possuindo baixa incidência nos chamados países ricos. Uma evidência dessa afirmação foi tornada pública quando a OMS publicou, em 2020, as dez maiores causas de morte no mundo (<https://www.who.int/data/global-health-estimates>), com dados obtidos a partir de informações epidemiológicas, separando os países em quatro grupos econômicos

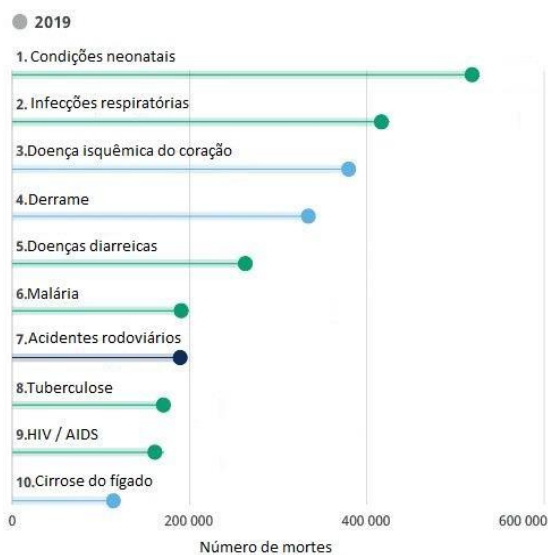
conforme propõe o Banco Mundial (2020). Ao compararmos as dez causas de morte entre os países mais ricos, com renda nacional bruta *per capita* (RNB) de U\$ 12.536,00 ou mais (Gráfico 2), com as de países de até U\$1.035,00 de RNB (Gráfico 3), podemos confirmar que as doenças negligenciadas, como a tuberculose, tem prevalência em países pobres e quase não existem entre as nações mais ricas (WHO, 2020).

Gráfico 2 - Causas de mortes em países de alta renda.



Fonte: Adaptado de <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> em 2019.

Gráfico 3 - Causas de morte em países de baixa renda.



Fonte: Adaptado de <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> em 2019.

O perfil de causas de mortes apresenta uma diferença significativa entre países ricos e pobres, conforme pode ser visto nos gráficos 2 e 3, sendo que uma característica desta diferença é a baixíssima produção de novos produtos terapêuticos para as denominadas doenças negligenciadas, nas quais inclui-se a TB (Pedrique *et al.* 2013). A tabela 1 apresenta os dados desta disparidade na geração de novos medicamentos.

Tabela 1 - Novos produtos terapêuticos aprovados ou recomendados por categoria de doença (2000 – 2011).

<i>Categoria</i>	<i>NEQ*</i> ( <i>n=336</i> )	<i>Outros **</i> <i>Novos Produtos</i> ( <i>n=420</i> )	<i>Vacinas *** ou</i> <i>Produtos</i> <i>Biológicos (n=94)</i>	<i>Total (n=850)</i>
Doenças negligenciadas	4 (1%)	25 (6%)	8 (9%)	37 (4%)
Outras doenças infecciosas	35 (10%)	48 (11%)	66 (70%)	149 (18%)
Demais doenças	297 (88%)	347 (83%)	20 (21%)	664 (78%)

\*Nova entidade química.

\*\* Novas indicações, formulações ou associações em dose fixa.

\*\*\*Inclui imunoglobulinas e outros produtos biológicos.

Fonte: Pedrique *et al.*, 2013.

De acordo com o trabalho de Pedrique *et al* (2013) apenas sete novos produtos terapêuticos (2000-2011) foram disponibilizados para o tratamento da TB (Tabela 2), sendo que nenhum deles desenvolvido em empresas de países de maior renda. Quanto aos ensaios clínicos, no mesmo período foram descritos apenas 474 (<0,5%) de um total de 148.445 estudos divulgados pelos *National Institutes of Health* e OMS.

Neste contexto de baixa produção científica, não podemos excluir a tuberculose multidroga resistente (TB-MDR), que é a TB simultaneamente resistente à isoniazida e à rifampicina (Brunton, 2012). A TB-MDR infecta 580 mil pessoas em todo planeta (BWHO, 2017), mas o financiamento em pesquisa e desenvolvimento (P&D) de novos medicamentos está entre os valores mais baixos desde 2008, totalizando um investimento de 620 milhões de dólares em 2015 (BWHO, 2017). A realidade do subfinanciamento em pesquisa assume maior

gravidade quando é apontado que a TB-MDR será responsável por 25% de todas as mortes por patógenos resistentes a fármacos (WHO, 2017). Em países como o Uzbequistão e Turcomenistão, a porcentagem da TB-MDR em relação às demais cepas, alcança até 30% dos casos (Golan, 2012).

Tabela 2 - Novos Produtos Terapêuticos para Tratamento da Tuberculose (2000-2011)

Produto Terapêutico	Ano de Aprovação	Titular de Comercialização
Ethambutol plus, isoniazid plus, pyrazinamide plus and rifampicin	2003	Lupin Ltd, Índia
Ethambutamol plus, isoniazid plus and rifampicin	2008	MacleodsPharmaceuticals Ltd, Índia
Isoniazidplus and rifampicin	2009	MacleodsPharmaceuticals Ltd, Índia
Isoniazid plus, pyrazinamide plus and rifampicin	2009	MacleodsPharmaceuticals Ltd, Índia
Moxifloxacin	2010	Cipla Ltd, Índia
Levofloxacin	2011	Cipla Ltd, Índia
Ofloxacin	2011	Cipla Ltd, Índia

Fonte: Adaptado de Pedrique *et al*, 2013

## 1.2 O BACILO *Mycobacterium tuberculosis*,

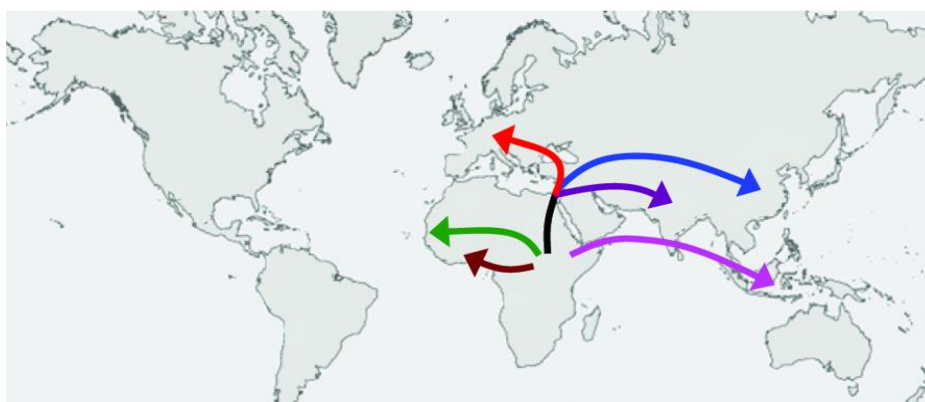
O agente causador da tuberculose foi isolado pela primeira vez em 1882, por Robert Koch (1843-1910), que o identificou como *Mycobacterium tuberculosis* (Lopes *et al*, 2016). O gênero *Mycobacterium* se caracteriza por conter bacilos aeróbicos, imóveis, não esporulados e não encapsulados. Estas bactérias possuem grandes teores de lipídios, sobretudo na parede celular, o que altera a permeabilidade destes microrganismos à água, a soluções corantes utilizadas em laboratório e a agentes desinfetantes (Lopes *et al*, 2016). Esse processo anabólico que mantém a parede celular é um dos alvos preferenciais para pesquisa de novos medicamentos, pois a formação desta estrutura, com aproximadamente 60% da

massa em ácidos graxos, além de não encontrar similaridade em mamíferos, representa fator de resistência a antibióticos e desinfetantes (Mota, 2011).

### 1.2.1 Origem e evolução

O estudo genético permite compreender o mecanismo de resistência da micobactéria, bem como sua origem e evolução. As pesquisas em um único locus, das várias cepas de TB, apontavam para uma baixa variabilidade da sequência de DNA (Brudey *et al*, 2006), entretanto trabalhos mais recentes, em uma abordagem multilocus (Baker *et al*, 2004), sugerem uma diversidade maior. O trabalho conduzido por Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, *et al*. (2008) propõe que, mais que a seleção natural, a deriva genética aleatória é uma grande responsável pela origem das diferentes cepas, sendo que desde o início da migração da humanidade, de seu berço na África rumo a outros continentes, a diferenciação do bacilo da TB ocorreu acompanhando os primeiros grupos do homem moderno. A figura 1 apresenta a diferenciação das cepas ao longo de, pelo menos, 50.000 anos de migrações humanas.

Figura 1 - Migração hipotética de linhagens ancestrais (preto, verde, marrom e lilás) e modernas (vermelho, azul e roxo) da micobactéria

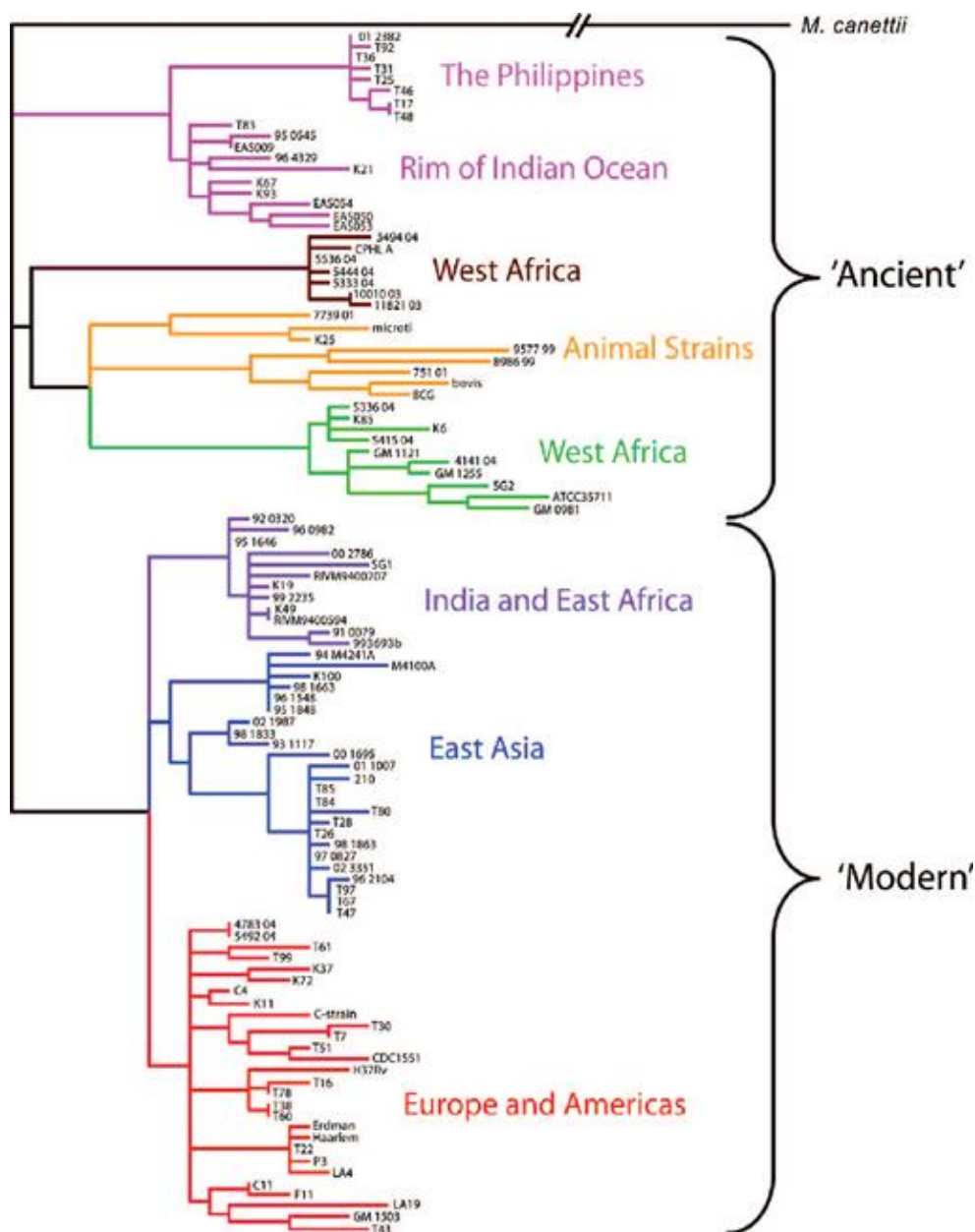


Fonte: Hershberg *et al* (2008)

As famílias da MTB podem ser divididas, em uma árvore filogenética, em cepas ancestrais e modernas com uma provável origem comum no bacilo *M. Canetti* (Fig. 2) sendo que, as cepas ancestrais teriam um tempo maior de latência

sem sintomas. Essa diferença faz sentido quando se pensa em populações humanas raras, dispersas e distantes umas das outras, já as cepas modernas, embarcando nos meios de transporte atuais e grande aumento populacional, evoluem para a forma ativa da tuberculose mais rapidamente (Hershberg *et al* 2008).

Figura 2 - Árvore filogenética do *Mycobacterium tuberculosis*.



Fonte: Hershberg *et al* (2008)

O estudo de Hershberg *et al* (2008) conclui que a maioria das mutações, cerca de 58%, ocorrem nas posições conservadas do DNA bacteriano, o que, além de outros indicadores, apontam uma redução da seleção natural, já que essas mutações poderiam ter um efeito levemente deletério nas micobactérias. No entanto, esse indicador de deriva genética aleatória com seleção reduzida, permitiria a persistência de cepas resistentes a medicamentos mesmo sem exposição a antibióticos e com um custo elevado para o bacilo. A pesquisa genética, conduzida por Hershberg e sua equipe, propõe que ocorre uma sinergia entre o MTB e as migrações humanas modernas, o que poderia ter um impacto tão importante, ou maior, que o uso inadequado de antibióticos no surgimento e manutenção de cepas resistentes a tratamento medicamentoso. Estudos mais recentes (Nguyen *et al*, 2018) apontam que a cepa selvagem de MTB pode desenvolver resistência a medicamentos, com alto ou baixo custo biológico, e se propagar por interferência clonal, sendo que, sob pressão medicamentosa podem ocorrer mutações compensatórias para aliviar o custo de adaptação destas mutações associadas a TB-MDR.

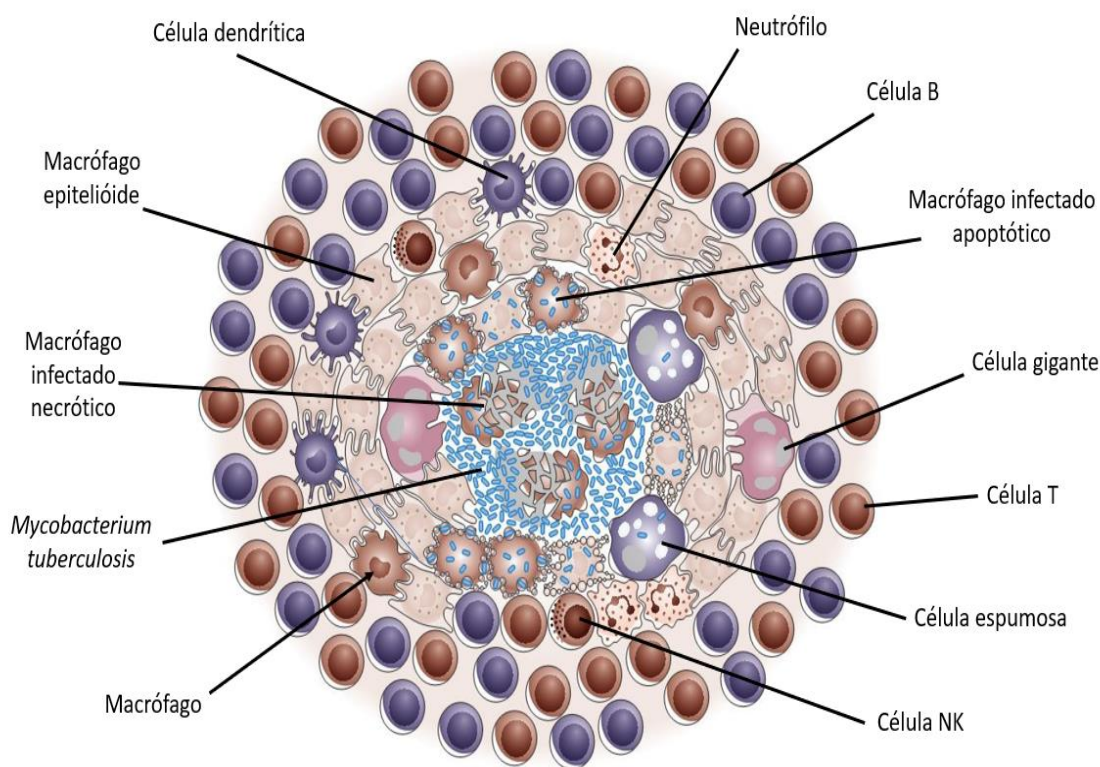
### 1.2.2 Patologia

A *Mycobacterium tuberculosis* infecta os humanos através das vias aéreas, depositando-se nos alvéolos pulmonares e, a seguir, é fagocitada pelos macrófagos. Infelizmente esta bactéria possui uma capacidade de escapar das diferentes estratégias do nosso sistema imune (Lopes *et al*, 2020). De acordo com Murray (2009), esta é uma etapa importante na etiologia da tuberculose, pois a micobactéria tem a habilidade de impedir a fusão do fagossoma com o lisossoma dentro do macrófago, passando o microrganismo a habitar o meio intracelular. Nesta condição, os macrófagos infectados passam a secretar interleucina-12 e o fator de necrose tumoral, desencadeando uma cascata de reações imunológicas, que podem ocasionar necrose tecidual.

Os macrófagos e outras células infectadas formam um núcleo necrosado, cercado de células T, CD4, CD8, NK e mais macrófagos, chamado granuloma (Ramakrishnan, 2012) (fig.3). Os maiores granulomas podem ser caseinosos e

encapsulados por fibrina, e estes protegem a micobactéria do sistema imune. A resposta imunológica, ao que tudo indica, é a única responsável pelos danos aos tecidos pulmonares. A produção de escarro sanguinolento indica a destruição tecidual e a evolução da patologia (Murray, 2009).

Figura 3 - Estrutura e constituintes do granuloma de *M. tuberculosis*

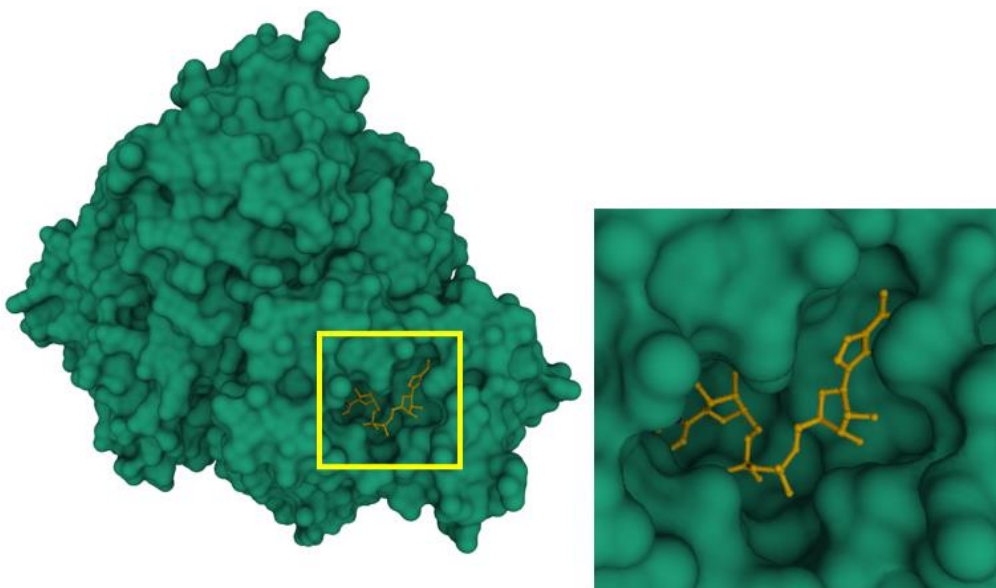


Fonte: Adaptado de Ramakrishnan (2012).

### 1.3 ALVO FARMACOLÓGICO E AÇÃO NO SÍTIO CATALÍTICO

Esta pesquisa tem como objetivo construir modelos matemáticos que possam apontar possíveis inibidores da enzima enoil-ACP-Redutase de *Mycobacterium tuberculosis* (InhA), que é o nosso alvo farmacológico. Esta enzima (Fig. 4) atua no ciclo FAS (*Fatty Acid Synthase*) no qual ocorrem reações repetitivas de condensação, ceto-redução, desidratação e enoil-redução para a síntese de ácidos graxos (AG) (Nelson, 2011).

Figura 4 - A enzima InhA, evidenciando em detalhe o cofator NDAH (amarelo) no sítio catalítico.



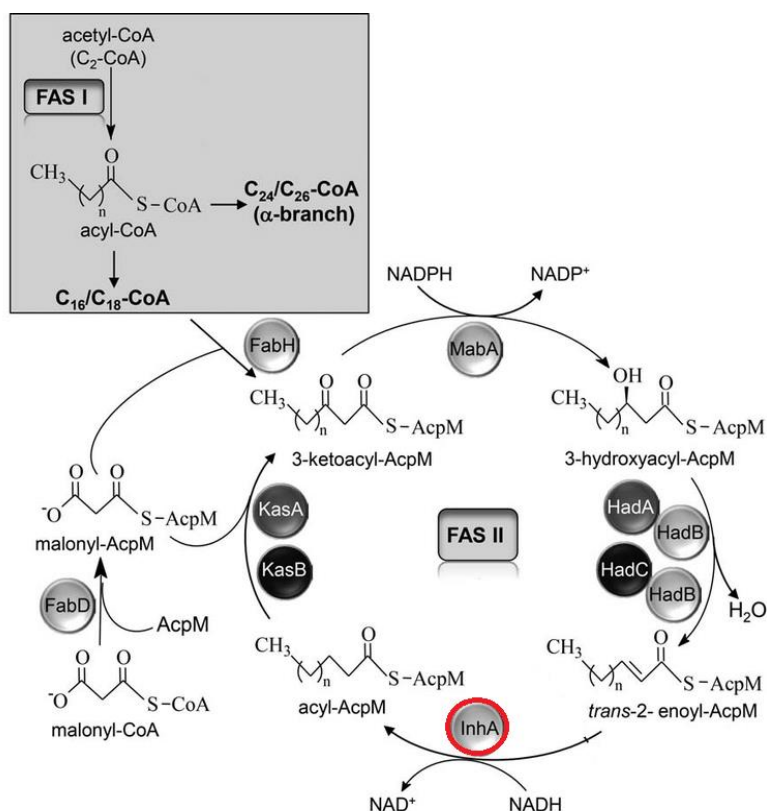
Fonte: Depositado na base de dados PDB sob código 6SQ5.

A MTB possui um envelope celular dinâmico e imunomodulador, que varia estruturalmente e funcionalmente ao longo do comprimento da célula, durante o processo de infecção (Dulberger *et al*, 2019). Esta variabilidade na parede celular da micobactéria pode permitir a adaptação ao meio, resistência ao sistema imune e tolerância ao tratamento com antibióticos que, no caso da MTB, é de 100 a 1000 vezes menos permeável aos fármacos  $\beta$ -lactâmicos do que as bactérias que não possuem parede celular (Gram-negativas) (Dulberger *et al*, 2019; Andrade, 2008). A estrutura serosa e rica em AG da parede celular exige uma cadeia de proteínas complexa e numerosa, com fatores de sinalização sensorial como quinases, transportadores, reguladores da parede celular e proteínas de montagem periplásmica (Dulberger *et al*, 2019).

Na maioria dos eucariotos o ciclo de síntese de AG, chamado FAS I, é executado por enzimas que trabalham agrupadas como uma grande enzima multifuncional (Nelson, 2011). Já nas bactérias, as enzimas sintetizam AG de forma diferente, em um ciclo chamado FAS II (Fig.5). As micobactérias possuem os dois ciclos, sendo o ciclo FAS II é responsável pela síntese dos ácidos micólicos (AM) (Duan *Xet al*, 2014) que compõem a parede celular do bacilo (Fig. 6) (Duan *Xet al*,

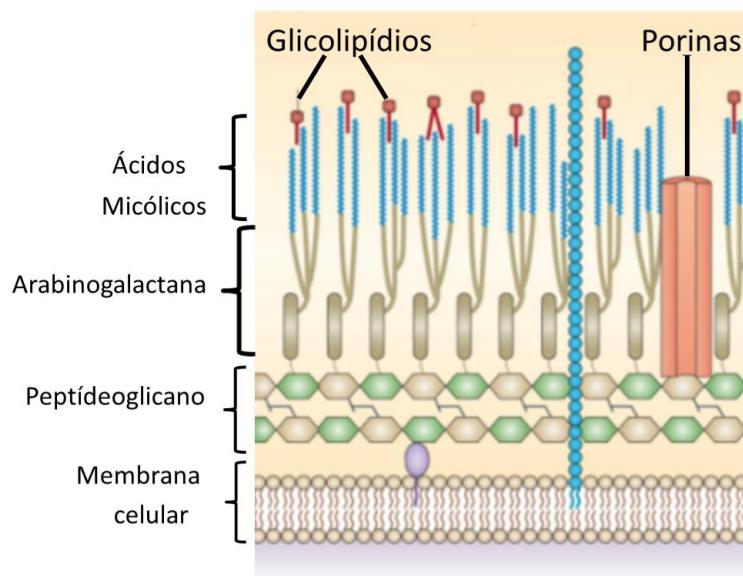
2014). Os ácidos micólicos consistem em ácidos graxos de cadeia longa 2-alkil, 3-hidroxila, nos quais o lipídio na posição 2, chamada de cadeia  $\alpha$ , está saturada e com 22 a 26 átomos de carbono de comprimento. O lipídio no meromicolato 3-hidroxila, tem 42 a 62 carbonos de comprimento. Já a cadeia  $\alpha$  e os primeiros carbonos do meromicolato são sintetizados pelas enzimas ácido graxo sintase I (FAS-I). As enzimas FAS-II então alongam essas cadeias nos meromicolatos promovendo modificações dos ácidos micólicos como dessaturações e ciclopropanações *cis* ou *trans* que dobram a cadeia de meromicolato sobre si mesma, bem como a adição de grupos hidroxila, ceto e metoxila orientado para a borda polar da membrana (Dulberger *et al*, 2019). Os AM ao final do processo FAS II são ácidos graxos longos, complexos e ramificados dos tipos alfa, metoxila e ceto sendo que os AM do tipo alfa formam aproximadamente 70% da composição de AM (Fig.7) (Brown *et al*,2015; Duan X *et al*, 2014;Dulberger *et al*, 2019).

Figura 5 - Ciclo FAS II. No destaque em círculo vermelho, a enzima InhA.

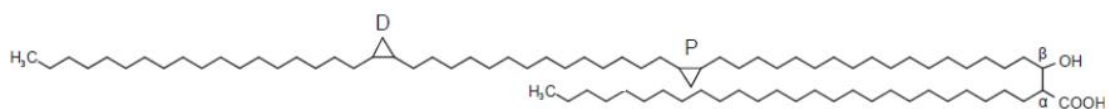


Fonte: Adaptado de Pawełczyk *et al*, 2014.

Figura 6 - Composição da parede celular da MTB



Fonte: Adaptado de Brown *et al*, 2015.

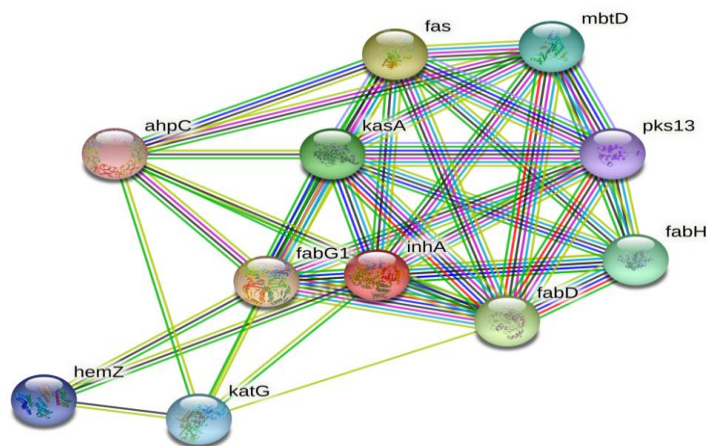
Figura 7 - Ácido micólico do tipo  $\alpha$ 

Fonte: Adaptado de Schiebel *et al*, 2013.

É desejável ter como alvo uma enzima que não encontre análogas em mamíferos e ainda possa evitar mecanismos de resistência, como os relatados em estudos com isoniazida e etionamida (Morlock *et al*, 2003; Duan X *et al*, 2014). Neste caso a InhA, uma redutase presente no ciclo FAS II, se apresenta como um objeto de pesquisa relevante (Freundlich JS *et al*, 2009; Khan *et al*, 2018; Morlock *et al*, 2003; Khade *et al*, 2020), pois além de não existir o ciclo metabólico FAS II em mamíferos, ela se relaciona com várias proteínas, como mostra a figura 8 gerada pela base de dados STRING (Szklarczyk *et al*, 2019). A rede de interação proteômica da InhA sugerida, é baseada em informações estabelecidas experimentalmente e em análises de genoma como: fusão de domínios, perfil filogenético e vizinhança genética (Raman, 2008; Szklarczyk *et al*, 2021).

Quando configurada para *full STRING network*, a base de dados aponta 11 proteínas como integrantes da denominada primeira camada de interação com alto escore de confiança estimada, e 21 proteínas em segunda camada de interação com escore menor. Entretanto, se for selecionada a opção de apontar as enzimas que fazem parte de um complexo físico, o número de proteínas reduz para 8 em primeira camada de interação e 18 em segunda camada (Szkłarczyk *et al*, 2021).

Figura 8 – Primeira camada de interações proteína-proteína conhecidas e previstas entre a InhA e outras enzimas.



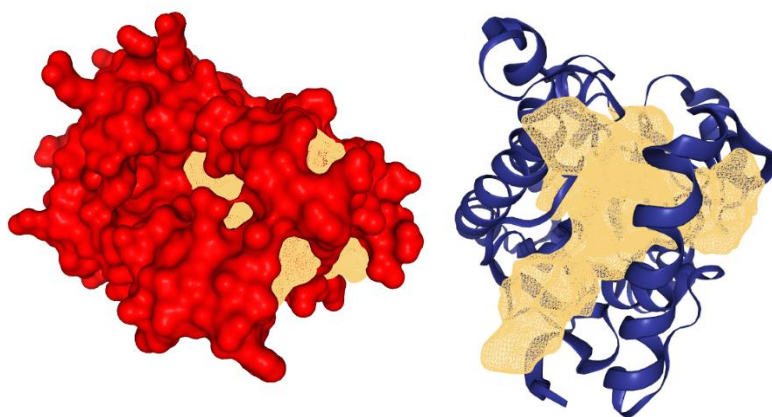
A cor do traço indica o tipo de evidência de interação: rosa – testado experimentalmente, verde – vizinhança genética, vermelho – fusão de genes, azul – co-ocorrência de genes, verde claro – *text mining*, preto – co-expressão e azul claro – homologia de proteínas. Obtido na base de dados STRING configurada em *full STRING network*, cepa H37Rv (Szkłarczyk *et al*, 2019).

Estas interações demonstram a complexidade do ciclo anabolizante de ácidos graxos para a formação da parede celular da micobactéria (Vilchèze e Jacobs, 2014; Morlock *et al*, 2003; Khade *et al*, 2020; Duan *et al*, 2014). A InhA possui um *binding pocket* relativamente volumoso (fig. 9) com volume estimado de 1.708,61 Å<sup>3</sup> e área de superfície de 2.124,91 Å<sup>2</sup> (Fährrolfes *et al*, 2017), permitindo a entrada do cofator NADH, substrato, moléculas inibidoras e/ou adutos, o que comporta várias ligações e interações moleculares, parte delas ainda não conhecida (Chollet *et al*, 2018).

Chollet *et al* (2018) propõem que o mecanismo de ação da InhA consiste na redução catalítica da ligação dupla *trans* do substrato do acil graxo, formando um intermediário enolato por meio da transferência do ânion hidreto do carbono na posição 4 da nicotinamida para a posição C3 do substrato (fig.10). Ao que parece, o resíduo Phe149 desempenha um papel crítico, para induzir a conformação

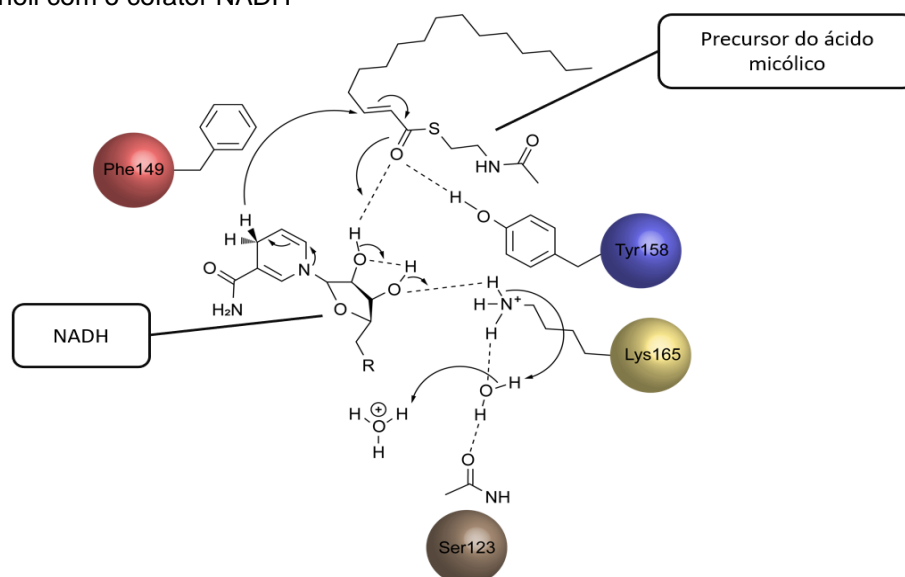
adequada da nicotinamida do cofator NADH. Estudos de mutagênese permitiram identificar Tyr158 e Lys165 como dois resíduos adicionais cataliticamente importantes. A estrutura do complexo InhA-NADH mostrou que a Lys165 estabelece ligação de hidrogênio com ambos os grupos hidroxila da porção ribosil ligada à nicotinamida (2,5 Å) (Chollet et al, 2018).

Figura 9 - *Pocket* do sitio catalítico da InhA.



Representação da enzima InhA nos formatos *molecular surface* (vermelho) e *cartoon* (azul). O volume do sítio catalítico está representado no formato malha (bege). Imagem gerada em <https://proteins.plus> (Fährrolfes et al, 2017).

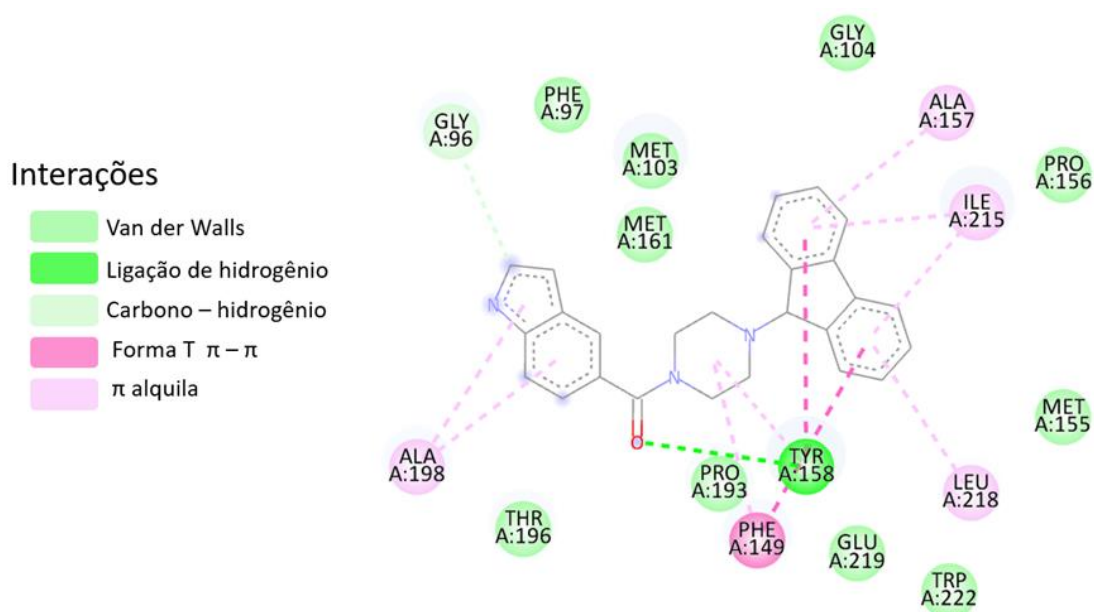
Figura 10 - Mecanismo proposto para redução catalisada pela enoil-ACP-redutase (InhA) do substrato enoil com o cofator NADH



Fonte: Adaptado de Chollet et al, 2018.

O resíduo Tyr158 aparentemente é um catalisador eletrofílico, estabilizando o intermediário enolato da reação. Uma vez que o enolato é formado, a transferência de um próton ocorre, envolvendo as hidroxilas da ribose e do resíduo Lys165, terminando em uma molécula de água (Chollet *et al*,2018). Já a Ser123 tem como papel permitir a transferência final do próton da molécula de água para a formação de um hidrônio. Esse processo todo permite a quebra de uma ligação  $\pi$  e o alongamento do futuro ácido micólico, encerrando a atividade enzimática da InhA, até receber novo substrato proveniente do agrupamento enzimático formado por HadA, HadB, HadC e HadB (Pawelczyk e Kremer, 2014; Chollet *et al*, 2018). Para corroborar a teoria da importância dos resíduos Lys165, Ser123, Tyr158 e Phe149 na ação catalítica, podemos observar que a localização do inibidor CID 447767 complexado a InhA (PDB 1P44) obtida por cristalografia de raio x, indica o posicionamento da arilamida no entorno destes resíduos vitais para a atividade biológica da enzima (Fig.11).

Figura 11 - Interações moleculares entre a molécula CID 447767 e a enzima InhA. Depositado sob código PDB 1P44



Fonte: Imagem do autor, gerada com BIOVIA DSV.

## 1.4 AÇÃO DE MEDICAMENTOS E MECANISMOS DE RESISTÊNCIA NO SÍTIO DE INTERESSE

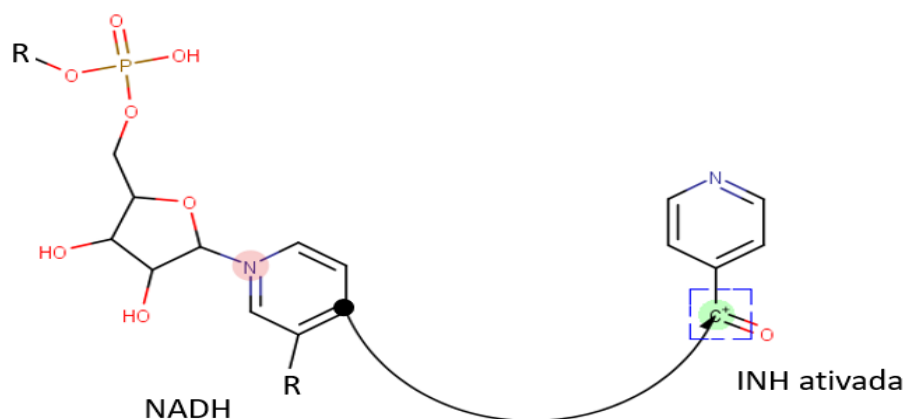
A isoniazida (INH) é considerada um fármaco vital no tratamento da TB, mas, infelizmente, o *Mycobacterium tuberculosis* evoluiu cepas resistentes a esta molécula (Vilchère e Jacobs, 2013; Lopes *et al*, 2020). Esse medicamento é um pró-fármaco ativado pela enzima KatG da MTB (Fig.12) sendo transformada em um radical que, ao reagir com o cofator  $\text{NAD}^+$ , forma um aduto covalente (Fig.13) (Lopes *et al*, 2020; Khadeet *al*, 2020). Esse aduto, por se ligar com alta afinidade ao sítio catalítico da enzima enoil-ACP-redutase (InhA), impede a utilização do cofator NADH para a produção dos ácidos micólicos que compõem a parede celular bacteriana (Morlock *et al*, 2003; Lopes *et al*, 2020; Khade *et al*, 2020), tendo como consequência a lise do bacilo da TB. A molécula etionamida (ETH) também é um pró-farmaco semelhante a INH (Fig. 14) mas sofre ativação através da enzima EthA e, como a INH, forma um aduto com o  $\text{NAD}^+$  (Fig. 15) (Morlock *et al*, 2003).

Figura 12 - Ativação do pró-fármaco INH via enzima KatG.



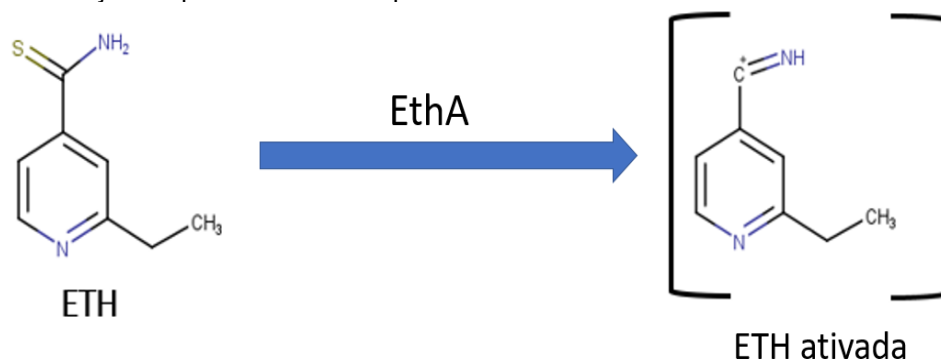
Fonte: Adaptado de Vilchère e Jacobs, 2014.

Figura 13 -Ligação da INH ativada ao cofator NAD<sup>+</sup> formando o aduto INH-NAD.



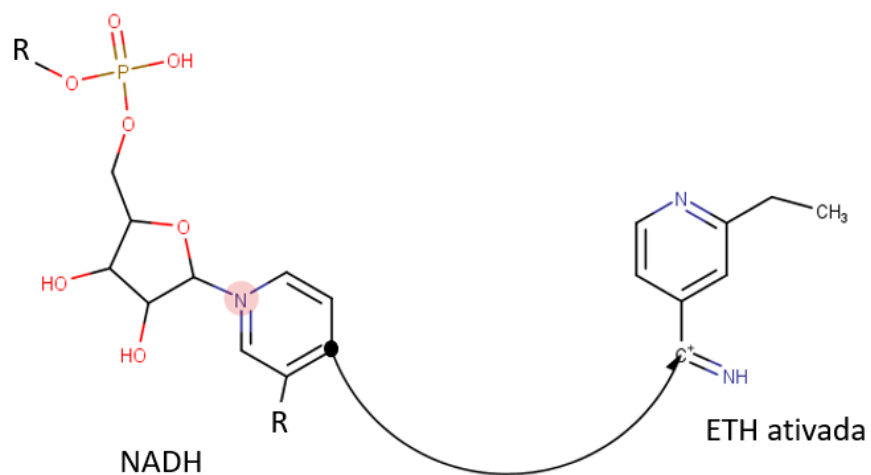
Fonte: Adaptado de Vilchêze e Jacobs, 2014.

Figura 14 - Ativação do pró-fármaco ETH pela enzima EthA.



Fonte: Adaptado de Vilchêze e Jacobs, 2014.

Figura 15 - Ligação da ETH ativada ao cofator NAD<sup>+</sup> formando o aduto ETH-NAD.

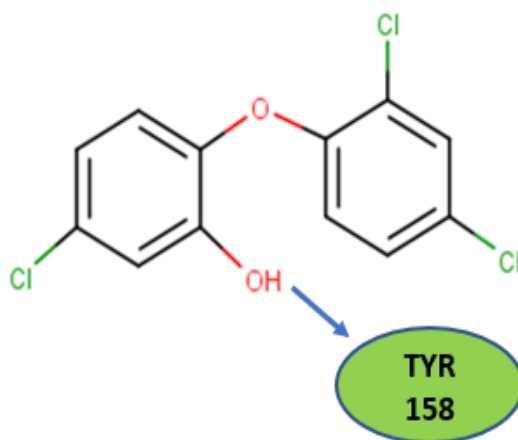


Fonte: Adaptado de Vilchêze e Jacobs, 2014.

Estudos apontam que mutações, principalmente no códon 315 do gene *KatG* (*KatG*), impedem a ativação da INH (Morlock *et al*, 2013), sendo que a mutação nesse gene é a causa da resistência à isoniazida em 64% dos isolados clínicos (Khade *et al*, 2020). Outras formas de resistência a INH foram observadas em mutações no gene estrutural e no promotor *InhA* (*InhA*), estas mutações reduzem a afinidade do NADH com a enzima *InhA* e promovem um mecanismo de titulação do medicamento (Vilchère e Jacobs, 2014; Morlock *et al*, 2003; Khade *et al*, 2020). Já no caso da resistência a etionamida, pesquisas apontam para mutações no gene *ethA* como responsável (Vilchère e Jacobs, 2014; Morlock *et al*, 2003; Khade *et al*, 2020). As mutações em *KatG* e *EthA* apontam a necessidade de buscar novos medicamentos cuja ação ocorra por inibição direta da enzima *InhA*, evitando etapas de ativação (Chollet *et al*, 2018).

Outro medicamento muito estudado para tratamento da tuberculose é o triclosan (TCL) que, ao contrário da INH e ETH, não precisa ser ativado para inibir a *InhA* (Khan *et al*, 2018), entretanto a sua biodisponibilidade é considerada subótima (Freundlich *et al*, 2009). O TCL se fixa ao sítio catalítico da *InhA* por intermédio de uma ligação de hidrogênio com o resíduo Tyr158, que representa 90% das interações entre TCL e a enzima (Khade *et al*, 2020) (fig. 16).

Figura 16 - Ligação de hidrogênio entre TCL e TYR 158



Fonte: Khade *et al*, 2020.

Um resumo dos estudos de mutagênese no sítio catalítico da InhA pode ser visto na tabela 3. Estas mutações podem conferir resistência a medicamentos ou até inibir a atividade enzimática.

Tabela 3 - Mutações de interesse em InhA.

Posição	Troca		Efeito
94	S	A	Confere resistência a INH e ETH. O mutante é 17 vezes mais resistente à inibição pelo aduto INH-NAD. Diminuição de 5 a 6 vezes na afinidade para NADH que resulta de uma perturbação na rede de ligações de hidrogênio que estabiliza a ligação de NADH. Quase nenhum efeito sobre a velocidade da enzima. Não tem impacto sobre a suscetibilidade à piridomicina
148	D	G	Confere resistência à piridomicina. Não tem impacto na suscetibilidade à isoniazida e moxifloxacina. Diminuição de 14 vezes na afinidade de NADH, sem efeito na atividade catalítica.
158	Y	A	Diminuição de 1500 vezes na atividade catalítica, sem efeito na afinidade do substrato lipídico.
158	Y	F	Diminuição de 24 vezes na atividade catalítica, sem efeito na afinidade do substrato lipídico.
158	Y	S	Nenhum efeito na atividade catalítica.
165	K	A	Perda da capacidade da enzima de se ligar ao NADH.
165	K	Q	Nenhum efeito na capacidade catalítica da enzima ou em sua capacidade de ligar NADH
266	T	A	Nenhum efeito na atividade catalítica. Perda de fosforilação. Não altera o crescimento de <i>M.tuberculosis</i> .
266	T	D/E	Prejudica gravemente a atividade catalítica, como consequência de uma afinidade de ligação reduzida ao NADH. Essas mutações de ponto único são letais para <i>M. tuberculosis</i> .

Fonte: Adaptado da base de dados Uniprot sob código P9WGR1.

## 1.5 O MODELO QSAR

Os estudos de QSAR (*Quantitative Structure-Activity Relationships*) têm como meta a construção de modelos matemáticos, que relacionam a estrutura química e a atividade biológica de uma série de compostos análogos (Ferreira *et*

al, 2002). A obtenção de uma correlação estrutura/atividade válida permite gerar uma função com capacidade preditiva quanto aos efeitos biológicos estimados de moléculas ainda não testadas *in vitro*, e que pode ser usada como ferramenta de busca de compostos promissores quanto ao resultado biológico desejado (Ferreira *et al*, 2002).

A ação terapêutica dos fármacos é resultante das interações destes com sistemas biológicos e é dependente de fatores relacionados com sua estrutura química e, conseqüentemente, de suas propriedades físico-químicas (Tavares, 2004). Estes fatores, sejam eles de caráter eletrônico, hidrofóbico ou estérico, influenciam a interação do fármaco com o alvo terapêutico e a sua distribuição nos diversos compartimentos que compõem o sistema biológico. Assim, fármacos estruturalmente semelhantes, podem apresentar diferenças relevantes quanto a sua atividade biológica mudando-se apenas um átomo na estrutura molecular, com implicações tanto quantitativas quanto qualitativas (Tavares, 2004).

A identificação de uma nova molécula com potencial terapêutico, na abordagem tradicional da química medicinal, requer elevado investimento em síntese orgânica, tempo e recursos financeiros. É fato que entre milhares de moléculas estudadas, poucas atingem os ensaios clínicos. Isso dificulta a descoberta de novas entidades químicas (NCEs) pelo processo tradicional (Narasimhan, 2014). Assim, vários métodos para determinar a atividade biológica de moléculas, sem a necessidade da síntese tradicional, têm sido pesquisados desde os anos 1960. Neste contexto, os estudos de QSAR buscam prestar agilidade, economia e confiabilidade na busca de NCEs (Narasimhan, 2014).

Um modelo matemático QSAR é desenvolvido com o objetivo de prever a resposta de estruturas semelhantes, sendo que o domínio de aplicação química é desenvolvido usando um grupo de análogos denominado conjunto de treinamento. A previsão de quaisquer novos produtos químicos, dentro do domínio, é considerada confiável se o mesmo passar pela validação interna (Roy, 2015). Já a chamada validação externa do modelo, é executada sobre moléculas não incluídas no desenvolvimento do modelo. Estas moléculas são denominadas Conjunto de Teste (Roy, 2015).

### 1.5.1 QSAR clássico, o modelo Hansh-Fujita

Hansch em 1960, identificou a dependência não linear (parabólica) da atividade biológica molecular com o seu coeficiente de partição (Log P) (NARASIMHAN, 2014) e propôs a seguinte função (1), onde (1/C) é a atividade biológica medida; Log P é o coeficiente de partição água-octanol; a, b, c são coeficientes de regressão:

$$\text{Log (1/C)} = a \log P - b (\log P^2) + c$$

Mais adiante, levando em consideração outros tipos de interações moleculares, Hansch e Fujita incluíram novos descritores, conhecidos como descritores moleculares, na função original para os efeitos estereoquímicos e eletrônicos, propondo o popular modelo Hansch e Fujita (clássico) (NARASIMHAN, 2014):

$$\text{Log (1/C)} = a (\text{desc. lipofílico}) + b (\text{desc. eletrônico}) + c (\text{desc. estérico}) + d (+ \text{descritores}) + e$$

O modelo clássico, utiliza a Regressão linear múltipla (MLR) onde as variáveis explicativas (independentes) procuram explicar a variável de resposta (dependente), neste caso os descritores moleculares (variáveis independentes) buscam estimar a atividade biológica (variável dependente).

### 1.5.2 QSAR 3D CoMFA

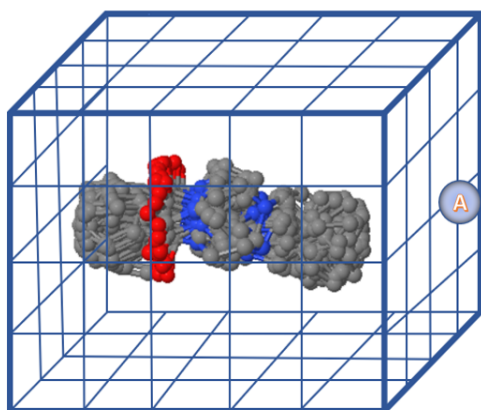
A metodologia tridimensional (QSAR-3D) tem se demonstrado valiosa para a pesquisa de novas moléculas com potencial farmacológico, e vários processos 3D neste sentido têm sido apresentados. Entretanto, o método de análise comparativa de campos moleculares (CoMFA) é, de longe, o mais utilizado e citado (Ragno, 2019).

Em 1988, Cramer *et al*, publicaram o artigo *Comparative Molecular Field Analysis (CoMFA). Effect of shape on binding of steroids to Carrier proteins*, nesse estudo foi apresentada a metodologia CoMFA como uma nova abordagem para a correlação estrutura/atividade. Suas principais características, propostas pela equipe de Cramer, são:

- representação de moléculas de ligantes por seus campos estérico e eletrostático, amostrados nas interseções de uma rede tridimensional,
- uma nova técnica de "ajuste de campo", permitindo o alinhamento mútuo ideal dentro de uma série, minimizando as diferenças de campo RMS entre as moléculas,
- análise de dados por mínimos quadrados parciais (PLS), usando a validação cruzada para maximizar a probabilidade de que os resultados tenham validade preditiva,
- representação gráfica dos resultados, como contornos energéticos tridimensionais e gráficos de coeficientes.

De acordo com a proposta de Cramer *et al* (1988), as moléculas submetidas ao modelo devem ser alinhadas em sua conformação de menor energia e centralizadas em um *grid* tridimensional, onde uma sonda virtual, simulando a aproximação de um átomo de carbono  $sp^3$  (ou outro átomo), vai calcular as energias eletrostáticas (Coulomb) e estéricas (potencial de Lennard-Jones) em cada ponto de intersecção do *grid*. Essa leitura vai permitir montar uma matriz de valores energéticos, que serão relacionados a atividade biológica por regressão de mínimos quadrados parciais (PLS) (fig.17) (Cramer *et al*,1988; Ragno, 2019; Cruz, 2012) gerando uma função preditiva e gráficos com a representação tridimensional dos campos energéticos estéricos e eletrostáticos.

Figura 17 - *GRID* 3D para cálculo de energias eletrostáticas e estéricas, onde as moléculas ficam centralizadas na conformação de menor energia global.



Molécula	Campo Eletrostático ( $\omega$ )	Campo Estérico ( $\Theta$ )
1	0,21334	0,84533
2	0,32466	0,7554
3	0,5600	0,2442
...	...	...

A sonda virtual, representado na imagem pela letra A, simula a aproximação de um átomo de carbono  $sp^3$  (ou outro) para cálculo das interações energéticas em cada cruzamento da malha do GRID entre a sonda e a molécula avaliada. Fonte: Imagem do autor.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Construir modelos matemáticos que estimam quantitativamente a atividade biológica de inibidores da enzima enoil-ACP-redutase de *Mycobacterium tuberculosis*.

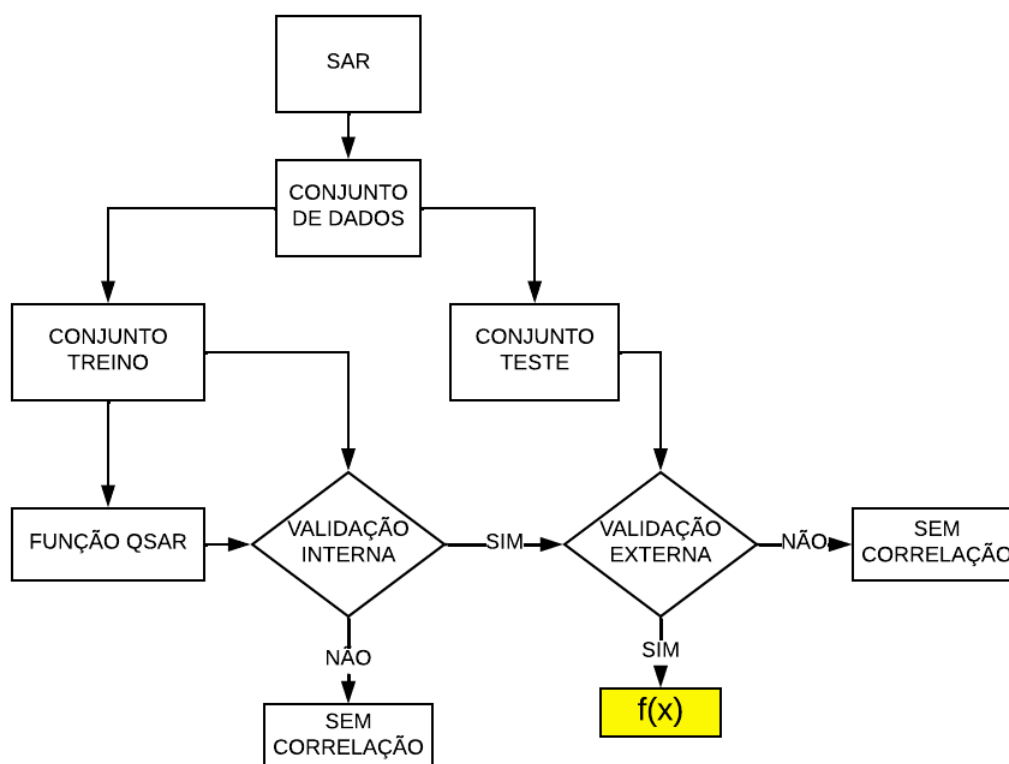
### 2.2 OBJETIVOS ESPECÍFICOS

- Identificar modelos QSAR para estimar a ação biológica de inibidores da enzima enoil-ACP-redutase de *Mycobacterium tuberculosis*.
- Estabelecer correlações entre resultados SAR, QSAR com docagem molecular, visando determinar o modelo matemático mais eficiente.
- Propor moléculas *drug-like* candidatas a etapa de ensaio *in vitro*.

### 3 METODOLOGIA

Os modelos QSAR clássico e CoMFA, seguiram o mesmo fluxograma de ações (Fig. 18).

Figura 18 - Fluxograma para construção dos modelos matemáticos QSAR clássico e CoMFA. Em amarelo, a função QSAR com capacidade preditiva obtida após a validação interna e externa.



Fonte: Imagem do autor.

#### 3.1 ESTUDO SAR

O estudo sobre a relação estrutura-atividade (SAR), promovido por He e colaboradores (2007), selecionou uma série de arilamidas (Fig.19) e avaliou *in vitro* a atividade biológica dos análogos sobre a enzima enoil ACP redutase de *Mycobacterium tuberculosis*. Da série pesquisada, 24 moléculas apresentaram atividade inibitória sobre a InhA (tabela 4) e serão utilizadas para a construção dos modelos matemáticos QSAR clássico e CoMFA, na forma de logaritmos dos valores

em micromolar de substância necessária para atingir 50% da concentração inibitória máxima (IC<sub>50</sub>).

Tabela 4 - Resultados do estudo de SAR promovido por He; Alian; Montellano, 2007.

CID*	IC <sub>50</sub> μM**	LogIC <sub>50</sub> μM ***
8366407	0,09	-1,045
447767	0,20	-0,699
9003925	0,40	-0,398
788974	0,99	-0,004
25093354	1,04	0,017
1012371	1,85	0,267
17312855	1,89	0,276
9378917	2,04	0,310
749123	3,07	0,487
1415427	5,16	0,713
1070143	6,05	0,781
807901	6,26	0,797
702506	6,73	0,828
1378917	7,39	0,869
814705	7,74	0,889
4006554	9,43	0,974
681767	9,74	0,989
739897	13,87	1,142
667749	14,11	1,150
2838469	15,47	1,189
749125	16,64	1,221
679013	17,62	1,246
116119	31,50	1,498
258240	38,86	1,590

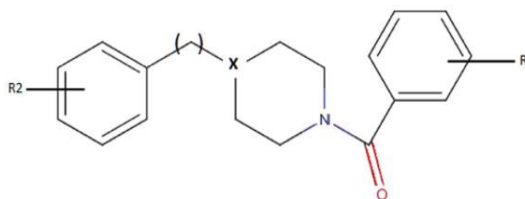
\* Código identificador da molécula (**CompoundID**) depositada na base de dados PubChem (S.Kim, 2016).

\*\* Atividade biológica medida em micromolar de substância necessária para atingir 50% da concentração inibitória máxima (IC<sub>50</sub>).

\*\*\* Atividade biológica calculada em logaritmo do valor em micromolar de IC<sub>50</sub>.

Fonte: HE *et al* (2007).

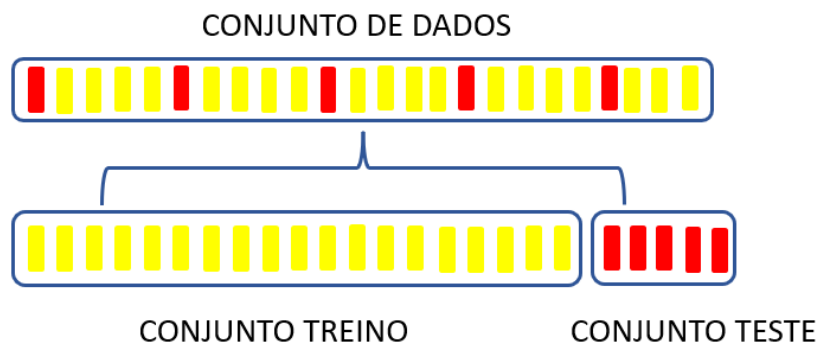
Figura 19 - Núcleo comum de uma série de análogos de arilamidas do estudo de SAR



Fonte: Adaptado de HE *et al* (2007).

De acordo com Martin *et al* (2012), todo conjunto de moléculas ativas descobertas durante o estudo de SAR, e que serão utilizados na modelagem QSAR, correspondem ao chamado *data set* e devem ser divididos em *training set* e *test set*, nas porcentagens de 80% para o primeiro e 20% para o segundo (fig. 20). As moléculas foram relacionadas em ordem decrescente de atividade biológica e selecionadas seguindo o critério de: uma molécula separada para o *test set*, as quatro subsequentes para o *training test*, outra molécula para o *test set*, mais quatro moléculas para o *training set* e assim por diante. A divisão das moléculas e seus respectivos códigos de identificação CID da base de dados PubChem (S.KIM *et al.* 2016), podem ser vistos na tabela 5. O conjunto de dados *training set* será utilizado para obter a função QSAR, com capacidade preditiva, na pesquisa de novas moléculas *drug-like*. A equação gerada é submetida à validação interna, onde se verifica se a função é capaz de, aproximadamente, reproduzir a atividade biológica das moléculas utilizadas para criar o próprio modelo, conforme o fluxograma de ações (fig.18). A seguir, na chamada validação externa, a função QSAR é testada em sua capacidade de estimar a atividade biológica do conjunto teste. Passando em ambas as validações, interna e externa, o modelo ainda é ranqueado por uma análise de qualidade dos dados preditivos obtidos.

Figura 20 - Divisão do conjunto de dados em conjunto treino e conjunto teste



Fonte: Imagem do autor.

### 3.2 FUNÇÃO QSAR CLÁSSICO

Todos os descritores moleculares usados neste modelo, foram calculados pelo programa MarvinView 17.24.0 (ChemAxon, 2017), e a geração do modelo matemático foi executada no programa BuildQSAR (Oliveira e Gaudio, 2000). O método escolhido para obter a função QSAR foi a *Multiple Linear Regression* (MLR). Trata-se de método eficiente quando se objetiva utilizar poucos descritores moleculares e número reduzido de compostos ( $\leq 30$ ) (Kutner *et al*, 2005). A validação interna foi obtida pelo programa BuildQSAR sendo que a análise de resíduos foi gerada, assim como a validação externa, para observar as características que comprometem a validade do MRL, de acordo com Amador *et al* (2011), utilizando a técnica de análise de regressão. Ao estabelecer um modelo de regressão, é necessário seguir alguns pressupostos que, conforme Levine *et al*. (2005), destacam-se os seguintes: homoscedasticidade (ou homogeneidade de variância), normalidade dos resíduos, independência dos erros e linearidade.

A análise estatística do modelo incluiu:

- a significância das variáveis independentes para o apontamento de atividade biológica,
- gráfico Q-Q,
- gráfico dos resíduos padronizados,

- gráfico de distância de Cook,
- coeficiente de correlação de Pearson  $R$ ,
- coeficiente de determinação  $R^2$ ,
- gráfico de dispersão,
- qualidade preditiva avaliada pelo indicador de relevância ou coeficiente de Stone-Geisser  $Q^2$ .

Em um modelo bem ajustado, isto é, que os erros do modelo são independentes e possuem distribuição Normal com uma mesma variância, esperamos que:

- a) Os pontos do gráfico dos valores ajustados x resíduos padronizados estejam em sua maioria entre -2 e 2 no eixo vertical e não tenham nenhum padrão de comportamento,
- b) que os pontos do QQ-plot estejam próximos da reta,
- c) que a densidade dos resíduos seja simétrica em forma de sino em torno do zero, variando de -2 a 2,
- d) Não haja pontos com alta Distância de Cook, que poderiam estar atrapalhando o ajuste do modelo.

O gráfico de correlação a ser gerado deve possuir uma reta pontilhada que representa a igualdade entre os valores observados e preditos, de forma que quanto mais próximo um ponto estiver dela, mais próximo estão os valores observados e preditos. Se o Intervalo de Predição intercepta essa reta, então temos que o modelo predisse o valor observado dentro da margem de erro. Além disso, geramos a linha sólida que representa a reta que melhor se ajusta aos resultados observados ( $y$ ) e os preditos ( $x$ ). Para completar a análise estatística também devem ser calculados os respectivos desvios padrão, fator de correlação de Pearson  $R$ , coeficiente de determinação  $R^2$  e o gráfico de dispersão (Neter *et al*, 1996; Kutner *et al*, 2005). Entretanto, como correlação não significa causalidade, os resultados devem ser submetidos ao indicador de Stone-Geisser  $Q^2$  para determinar a qualidade preditiva (Chin, 2010).

Tabela 5 - O conjunto de moléculas que compõem o estudo de SAR, divididas em *test set* e *training set*, e identificadas pelo código CID utilizado na base dados PubChem.

Molécula	<i>Test set</i> (CID)	<i>Training set</i> (CID)
1	8366407	
2		447767
3		9003925
4		788974
5		25093354
6	1012371	
7		17312855
8		9378917
9		749123
10		1415427
11	1070143	
12		807901
13		702506
14		1378917
15		814705
16	4006554	
17		681767
18		739897
19		667749
20		2838469
21	749125	
22		679013
23		116119
24		258240

Moléculas ordenadas por ordem decrescente de atividade biológica, conforme tabela 4.

### 3.3 CoMFA

Para a modelagem CoMFA foi utilizado a ferramenta desenvolvida pelo Prof. Rino Ragno (Ragno, 2019). Assim como no modelo QSAR clássico, foi utilizada a mesma distribuição das moléculas do estudo de SAR, divididas *em training set* e *test set*. As validações interna e externa foram executadas pela ferramenta do professor Ragno (2019) bem como a geração de gráficos 3D dos campos energéticos. No gráfico a ser gerado, a reta pontilhada representa a igualdade entre os valores observados e preditos, de forma que quanto mais próximo um ponto estiver dela, mais próximo estão os valores observados e preditos. Além disso, temos a reta sólida que representa a reta que melhor se ajusta aos pontos do gráfico, estabelecendo uma relação entre os valores do experimento (y) e os preditos (x). A análise estatística foi feita para obter os respectivos desvios padrão, fator de correlação de Pearson  $R$ , coeficiente de determinação  $R^2$  e o gráfico de dispersão (Neter *et al*, 1996), além da qualidade preditiva via indicador de Stone-Geisser  $Q^2$  (Chin, 2010).

#### 3.3.1 Alinhamento

O alinhamento das moléculas utilizou o método RDkit, tendo como referência o composto mais ativo do *training set* (CID447767), na conformação de menor energia global.

#### 3.3.2 Geração do modelo

A formatação para busca das conformações e geração do modelo seguiu as seguintes configurações:

- Sonda átomo de carbono  $Sp^3$  carga +1,
- Grid com extensão de 8 Å,
- Constante dielétrica de 18,

- *Cutoff* de 20 kcal/mol,
- Teste externo: *true*.

### 3.4 VIRTUAL SCREENING

O processo de busca de moléculas com o mesmo núcleo das arilamidas, utilizou como ferramenta de seleção o coeficiente de Tanimoto, disponível na base de dados pública PubChem (Kim *et al*, 2016). Dos compostos análogos encontrados, foram baixados os respectivos arquivos padrão SDF para a segunda etapa do processo, que é calcular os descritores moleculares dos compostos exigidos pela função QSAR. Para isto é utilizado novamente o programa MarvinView17.24.0 (ChemAxon, 2017). Na sequência, os resultados dos DM são submetidos a função QSAR para cálculo da atividade biológica, ordenando os resultados em ordem decrescente de Log IC<sub>50</sub>. Finda esta etapa, as 10 moléculas com maior atividade biológica estimada foram submetidas a fase de docagem molecular.

### 3.5 DOCAGEM MOLECULAR

Para esta etapa foi utilizada a primeira molécula do conjunto treino (CID 447767), devido a sua elevada atividade biológica observada no estudo de He e equipe (2007). A seguir foi pesquisado na base de dados *Protein Data Bank* (PDB) (Berman *et al*, 2000) uma estrutura que apresentasse a molécula CID 447767 ligada a InhA. Encontrado o complexo desejado, o arquivo formato PDB obtido foi editado de forma que pudesse ser trabalhado no programa PyRx Python Prescription 0.8 (Dallakyan, 2015). A seguir, estas moléculas são submetidas ao *Molecular docking* via programa PyRx utilizando AutoDock Vina (Trott e Olson, 2010). Os resultados são classificados conforme a energia de ligação EBEF (*Estimated Free Energy of Binding*). A imagem 3D do complexo obtido na docagem molecular é gerada com o aplicativo PyMol (The PyMOL Molecular Graphics System, Version1.8, 2015) e as interações energéticas foram processadas

utilizando os programas BIOVIA DSV (Discovery Studio Visualizer v21.1.0.20298) e BINDSURF (Sánchez-Linares, Irene *et al*, 2012).

## 4 RESULTADOS

Os resultados da presente dissertação serão apresentados na forma de artigo a ser submetido à revista *Journal of Computer-Aided Molecular Design*.

---

### Desenvolvimento de Modelos QSAR para Identificar Inibidores da Enzima enoil-ACP-redutase de *Mycobacterium tuberculosis*

Aureo André Karolczak<sup>a</sup>; Luis Fernando S. M. Timmers<sup>b</sup>; Rafael Andrade Caceres<sup>c</sup>

<sup>a</sup> Programa de Pós Graduação em Biociências da UFCSPA – Universidade Federal de Ciências da Saúde de Porto Alegre. Porto Alegre/RS, CEP 90050-170, Brasil. Contato: aureoa@ufcspa.edu.br

<sup>b</sup> Programa de Pós Graduação em Biotecnologia da UNIVATES – Universidade do Vale do Taquari. Lageado/RS CEP 95914-014, Brasil. Contato: luis.timmers@univates.br

<sup>c</sup> Programa de Pós Graduação em Biociências da UFCSPA – Universidade Federal de Ciências da Saúde de Porto Alegre. Porto Alegre/RS, CEP 90050-170, Brasil. Contato: rafaelca@ufcspa.edu.br

---

#### RESUMO

A tuberculose é motivo de preocupação mundial devido a sua prevalência em países em desenvolvimento, e pela capacidade da micobactéria de apresentar resistência aos esquemas terapêuticos existentes. Nesse contexto, o presente projeto se propõe a modelagem QSAR (*Quantitative Structure-Activity Relationships*), como forma de identificar e avaliar, *in silico*, a atividade inibitória estimada de moléculas candidatas às etapas de aperfeiçoamento molecular e/ou ensaios *in vitro*, reduzindo tempo de pesquisa e custos financeiros. Para tanto, foi utilizado o estudo de SAR (*Structure-Activity Relationships*) conduzido por He, Alian e Montellano (2007) sobre uma série de arilamidas, testadas como inibidoras da enzima enoil-ACP-redutase (InhA) de *Mycobacterium tuberculosis*. Foram desenvolvidos os modelos QSAR Hansh-Fujita (clássico) e CoMFA (*Comparative Molecular Field Analysis*). O modelo QSAR clássico obteve o melhor resultado estatístico utilizando *Multiple Linear Regression* (MLR), sendo a validação interna com fator de correlação  $R^2 = 0,9012$  e qualidade preditiva, conforme o indicador de Stone-Geisser  $Q^2 = 0,8612$ . Já na validação externa foi obtido um fator de correlação  $R^2 = 0,9298$  e um  $Q^2 = 0,720$ , indicando um modelo matemático altamente preditivo. O Modelo CoMFA logrou obter um  $Q^2 = 0,6520$  na validação interna, o que permitiu estimar os campos energéticos no entorno das moléculas utilizadas, informação essencial para propor aperfeiçoamento molecular. Construiu-se uma biblioteca de pequenas moléculas, com análogos aos utilizados no estudo de SAR, que foram submetidos a função QSAR clássico, tendo como resultado um grupo de dez moléculas com elevada atividade biológica estimada. Os resultados de docagem molecular sugerem que os dez análogos apontados pelo modelo QSAR clássico apresentaram energia livre de ligação estimada favorável. A conclusão aponta a metodologia QSAR como uma ferramenta eficiente e eficaz na busca e identificação de moléculas *drug-like* promissoras.

Palavras-chave: Tuberculose. QSAR. CoMFA. InhA. Docagem.

---

## 1 Introdução

De acordo com o Relatório Global da Tuberculose 2020, publicado pela Organização Mundial da Saúde (OMS), a doença causada pelo *Mycobacterium tuberculosis* (MTB) está entre as mais letais da atualidade, sendo classificada como uma das 10 maiores causas de mortes no mundo e a principal causa de morte tendo um único agente de infecção. Outro dado relevante é a informação que aproximadamente um quarto da população mundial está infectado com o MTB. O Estudo conduzido pela OMS aponta uma correlação entre a pobreza com a incidência da enfermidade sugerindo que, quanto maior a renda *per capita* menor será a incidência da tuberculose (TB), deixando claro o aspecto social que envolve essa doença [1].

Uma característica perversa desta diferença social é o baixíssimo financiamento e produção de novos produtos terapêuticos para as conhecidas doenças negligenciadas, na qual se inclui a TB [2]. O contexto de subfinanciamento em pesquisa assume maior gravidade quando é projetado que a tuberculose multidroga resistente (TB-MDR) será a responsável por 25% de todas as mortes por patógenos resistentes a drogas [3]. Em países como o Uzbequistão e Turcomenistão, a porcentagem da TB-MDR em relação às demais cepas alcançam até 30% dos casos [4].

### 1.1. Modelos QSAR

Uma abordagem útil para acelerar estudos de novos medicamentos e diminuir custos de pesquisa é a utilização da modelagem QSAR (*Quantitative Structure-Activity Relationships*) para a construção de modelos matemáticos que relacionem a estrutura química com a atividade biológica de uma série de compostos análogos [5]. Estes modelos são gerados a partir de estudos de SAR (*Structure-Activity Relationships*) anteriores. Os modelos QSAR selecionados nesta pesquisa foram:

#### 1.1.1. Hansh-Fujita ou clássico

Busca relacionar a atividade biológica com os chamados descritores moleculares (DM), por meio da regressão linear múltipla (RLM) [6]. Neste caso, os descritores moleculares (variáveis independentes) buscam prever a atividade biológica (variável dependente).

#### 1.1.2. CoMFA

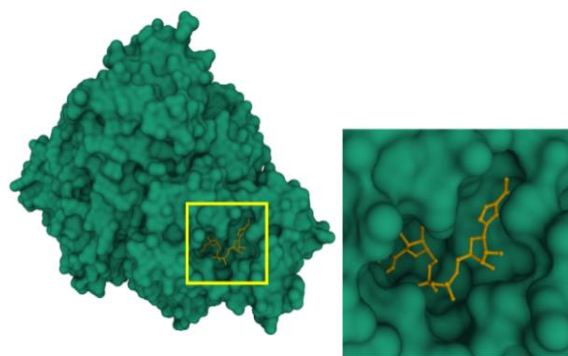
O modelo CoMFA (*Comparative Molecular Field Analysis*), conforme Cramer e equipe [7], utiliza a representação de moléculas de ligantes por seus campos de interação energéticos estéricos e eletrostáticos, amostrados nas interseções de uma rede

tridimensional (GRID). Também usa uma técnica de "ajuste de campo" que permite o alinhamento mútuo ideal dentro de uma série, minimizando as diferenças de campo RMS entre as moléculas e faz a análise de dados por mínimos quadrados parciais (PLS). O modelo permite a representação gráfica dos resultados, como gráficos de coeficientes e contornos energéticos tridimensionais [7].

### 1.2. Alvo farmacológico

Esta pesquisa tem como objetivo construir modelos matemáticos que possam apontar possíveis inibidores da enzima enoil-ACP-redutase (InhA), de *Mycobacterium tuberculosis*, que será o nosso alvo farmacológico. Esta enzima (Figura 1) atua no ciclo FAS II (*Fatty Acid Synthase*) no qual ocorrem reações repetitivas de condensação, ceto-redução, desidratação e enoil-redução para a síntese de ácidos graxos (AG) [8]. Estudos apontam que a ação enzimática ocorre em torno dos resíduos de aminoácidos Tyr158, Lys165, Ser123 e Phe149, no sítio catalítico [9].

**Figura 1.** A enzima InhA evidenciando, no detalhe, o cofator NADH (amarelo) no sítio catalítico.



Depositado na base de dados Protein Data Bank (PDB) sob código 6SQ5.

É desejável ter como alvo uma enzima a qual não são encontradas análogas entre mamíferos e que possa evitar mecanismos de resistência, como os relatados em estudos com isoniazida e etionamida [10]. Neste caso a InhA se apresenta como um objeto de pesquisa relevante [10-13].

## 2 Metodologia

### 2.1 Fluxograma dos modelos QSAR

Os modelos QSAR clássico e CoMFA, seguiram o mesmo fluxograma de ações (Figura 2).

**Figura2.** Fluxograma para construção dos modelos matemáticos QSAR clássico e CoMFA. Em amarelo, a função QSAR com capacidade preditiva obtida após a validação interna e externa.

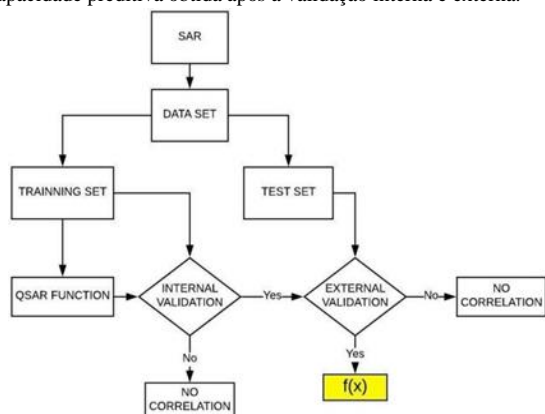
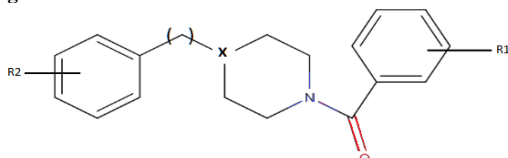


Imagem do autor.

## 2.2. Estudo de SAR

Em 2007, He, Alian e Montellano [16] fizeram um estudo de inibição, tendo como alvo a InhA de micobactérias. Para tanto foi utilizado uma série de arilamidas (Fig. 5). O bioensaio reportou 24 moléculas ativas (Apêndice II) que estão descritas na base de dados pública PubChem [17] sob código AID 301095.

**Figura5.** Estrutura Básica da arilamida.



Fonte: He, Alian e Montellano (figura modificada) [16]

## 2.3. Classificação dos dados

De acordo com Martin e colaboradores [18] todo conjunto de moléculas ativas, utilizadas para modelos QSAR, correspondem ao chamado *data set* e devem ser divididos em *training set* e *test set* nas porcentagens de 80% para o primeiro e 20% para o segundo (Figura 6) [19]. Neste estudo utilizamos uma proposta adaptada de Damre e colaboradores [20] que leva em consideração a diversidade estrutural e os valores de atividade biológica.

**Figura 6.** Divisão das moléculas em conjunto treino e conjunto teste, de acordo com Martín *et al* [18] e Leonard e Roy [19].



. Fonte: Imagem do autor

A proposta adaptada de Damre e equipe [20], envolve 2 passos:

a) colocar os compostos em ordem decrescente de valores de atividade biológica medida ( $IC_{50}$ ,  $EC_{50}$ , MIC

ou  $K_i$ ) na forma de logaritmo. Neste estudo optou-se pelo índice  $Log IC_{50}$  que é o logaritmo da metade da concentração de substância necessária para atingir a atividade inibitória máxima.

b) transferir o primeiro composto para o *test set*, separar os 4 ou 5 compostos seguintes para o *training set* [18], o composto subsequente é destinado para o *Test set*. O processo repete até que todos os compostos que compõe o *data set* estejam divididos.

## 2.4. Modelo Clássico

### 2.4.1. Descritores moleculares

Para o QSAR clássico foram avaliados 18 descritores moleculares. A maioria calculados com o uso do programa MarvinView 17.24.0[21], utilizando arquivos formato SDF obtidos da base de dados Pubchem [17]. A solubilidade em água foi obtida utilizando o aplicativo web ALOGPS 2.1 [22] utilizando código *smiles* da base de dados PubChem. Os DM utilizados estão especificados no Apêndice I.

### 2.4.2 Função QSAR

Com o *training set* definido, todos os descritores moleculares calculados e a atividade biológica ( $LogIC_{50}$ ) obtida do estudo de SAR, os dados foram processados utilizando o programa *BuildQSAR* [23] para obter a equação QSAR. O método escolhido foi a regressão linear múltipla (RLM), trata-se de método eficiente quando se objetiva utilizar poucos descritores moleculares e número reduzido de compostos ( $\leq 30$ ) [20]. O processo envolveu avaliar a relação de cada descritor molecular individualmente com a atividade biológica, depois realizar o mesmo processo utilizando pares de DM e, por fim, usando 3 descritores moleculares para correlacionar com o  $LogIC_{50}$ . Foram obtidas combinações com bons resultados, porém optou-se por escolher os descritores que alcançaram a maior correlação sem gerar *outliers* na validação interna.

### 2.4.3 Validação interna

A validação interna da função QSAR utilizou a análise dos resíduos para avaliar o ajuste de dados e interpretar o modelo RLM obtido com o *training set*. Em um modelo bem ajustado, isto é, que os erros do modelo são independentes e possuem distribuição Normal com uma mesma variância, esperamos que:

- Os pontos do gráfico dos valores ajustados x resíduos padronizados estejam em sua maioria entre -2 e 2 no eixo vertical e não tenham nenhum padrão de comportamento,
- que os pontos do QQ-plot estejam próximos da reta,
- que a densidade dos resíduos seja simétrica em forma de sino em torno do zero, variando de -2 a 2,
- Não haja pontos com alta Distância de Cook, que poderiam estar atrapalhando o ajuste do modelo.

O gráfico de correlação a ser gerado deve possuir uma reta pontilhada que representa a igualdade entre os valores observados e preditos, de forma que quanto mais próximo um ponto estiver dela, mais próximo estão os valores observados e preditos. Se o Intervalo

de Predição intercepta essa reta, então temos que o modelo predisse o valor observado dentro da margem de erro. Além disso, temos a linha sólida que representa a reta que melhor se ajusta aos resultados observados ( $y$ ) e os preditos ( $x$ ). Para completar a análise estatística também devem ser calculados os respectivos desvios padrão, fator de correlação de Pearson  $R$ , coeficiente de determinação  $R^2$  e o gráfico de dispersão [24]. Entretanto, como correlação não significa causalidade, os resultados devem ser submetidos ao indicador de Stone-Geisser  $Q^2$  para determinar a qualidade preditiva [25].

#### 2.4.4 Validação externa

O *test set* composto de 5 moléculas [20] foi submetido a função QSAR onde, a atividade biológica foi estimada. Os resultados calculados foram comparados com a atividade biológica observada no bioensaio AID 301095, para obter os respectivos desvios padrão e os fatores de correlação de Pearson  $R$ , coeficiente de determinação  $R^2$  e o gráfico de dispersão [24] além da qualidade preditiva via indicador de Stone-Geisser  $Q^2$  [25].

#### 2.4.5 Triagem virtual

O processo de busca de moléculas com o mesmo núcleo das arilamidas, utilizou como ferramenta de seleção o coeficiente de Tanimoto, disponível na base de dados pública PubChem [17]. Encontrados os compostos análogos, foram baixados os respectivos arquivos padrão SDF para a segunda etapa do processo que é calcular os descritores moleculares dos compostos, exigidos pela função QSAR. Para isto é utilizado novamente o programa MarvinView17.24.0 [21]. Na sequência, os resultados dos DM são submetidos a função QSAR para cálculo da atividade biológica, ordenando os resultados em ordem decrescente de Log IC<sub>50</sub>. Finalizada a etapa, as 10 moléculas com maior atividade biológica estimada passaram a fase de docagem molecular.

#### 2.4.6. Docagem molecular

Para essa etapa foi utilizada a molécula 2 do conjunto treino (CID 447767) devido a sua elevada atividade biológica *in vitro*, observada no bioensaio AID 301095. A seguir foi pesquisado, na base de dados *Protein Data Bank* (PDB) [26], uma estrutura que apresentasse a molécula 2 ligada a InhA. Encontrado o complexo desejado, o arquivo formato PDB obtido foi editado para que fosse compatível para o programa PyRx Python Prescription 0.8 [27]. A seguir todas as 10 moléculas selecionadas na triagem virtual foram submetidas ao *Molecular docking* via programa PyRx com AutoDock Vina. Os resultados são classificados conforme a energia de ligação EFEB (*Estimated Free Energy of Binding*) sendo que a molécula com a menor *binding energy* é comparada ao complexo depositado na base de dados PDB usado como modelo. A imagem 3D do novo complexo obtido via docagem molecular é gerada com o aplicativo PyMol (The PyMOL Molecular Graphics System, Version1.8, 2015) e suas

interações moleculares no sítio catalítico são obtidas com o programa BIOVIA *Discovery Studio Visualizer* [28]. As Contribuições para a energia global de ligação ao sítio catalítico foram calculadas com o *Blind Docking Server*, disponível em: <<http://bio-hpc.eu/software/blind-docking-server/>> [29].

### 2.5 Modelo CoMFA

Para a modelagem CoMFA foi utilizado a ferramenta desenvolvida pelo Prof. Rino Ragno [30]. Assim como no modelo QSAR clássico, foi utilizada a mesma distribuição das moléculas do estudo de SAR, divididas em *training set* e *test set*. As validações interna e externa foram executadas pela ferramenta do professor Ragno, bem como a geração de gráficos 3D dos campos energéticos. No gráfico a ser gerado, a reta pontilhada representa a igualdade entre os valores observados e preditos, de forma que quanto mais próximo um ponto estiver dela, mais próximo estão os valores observados e preditos. Além disso, temos a reta sólida que representa a reta que melhor se ajusta aos pontos do gráfico, estabelecendo uma relação entre os valores do experimento ( $y$ ) e os preditos ( $x$ ). A análise estatística foi feita para obter os respectivos desvios padrão, fator de correlação de Pearson  $R$ , coeficiente de determinação  $R^2$  e o gráfico de dispersão [24], além da qualidade preditiva via indicador de Stone-Geisser  $Q^2$  [25].

#### 2.5.1 Alinhamento

Para o alinhamento das moléculas foi utilizado o método RDKit, tendo como referência o composto mais ativo do *training set* (CID447767), na conformação de menor energia global.

#### 2.5.2 Geração do modelo

A formatação para busca das conformações e geração do modelo seguiu as seguintes configurações:

- Sonda átomo de carbono Sp<sup>3</sup> carga +1,
- Grid com extensão de 8Å,
- Constante dielétrica de 18,
- *Cutoff* de 20 kcal/mol,
- Teste externo: *true*.

## 3 Resultados e discussão

### 3.1 Divisão dos dados para ambos modelos

Foram selecionados para o *training set* as moléculas classificadas como: 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15, 17, 18, 19, 20, 22, 23 e 24, na ordem decrescente de atividade biológica inibitória. Para o *test set* as moléculas escolhidas foram as classificadas como: 1, 6, 11, 16 e 21. A identificação de cada molécula e a atividade inibitória, descritas pelo bioensaio AID 301095, podem ser conferidos no Apêndice III.

### 3.2. Modelo clássico

#### 3.2.1. Cálculo dos descritores moleculares

Ao todo, 18 descritores moleculares (Apêndice I) foram calculados e avaliados quanto a sua correlação com a atividade inibitória das moléculas que compõem o conjunto treino. Foram selecionados para compor a

função QSAR, os seguintes DM que obtiveram a máxima correlação sem produzir valores *outliers*:

- HA – número de aceptores de hidrogênio,
- RB – número de ligações rotáveis,
- VWALLS V – volume de Van der Walls.

O apêndice IV contém a Tabela com os valores calculados, para os 3 descritores moleculares escolhidos relacionados com as 19 moléculas que compõem o *training set*.

### 3.2.2 Função QSAR

Com os valores dos 3 DM calculados (Apêndice IV) mais a atividade inibitória (Log IC<sub>50</sub>) das moléculas que compõem o *training set*, foi obtida a função preditiva que pode ser vista na equação (Equação 1) que representa a função QSAR para a série de moléculas avaliadas por He, Alian e Montellano [16].

#### Equação 1. Função QSAR.

$$\text{LogIC}_{50} = 0,1727 \text{ HA} + 0,3886 \text{ RB} - 0,0204 \text{ VWALLS V} + 5,4255$$

### 3.2.3 Validação interna

Após obter a atividade biológica calculada com o uso da função QSAR (Tabela 3), os resultados foram avaliados estatisticamente quanto a adequação dos dados utilizados e comprovou-se que os parâmetros das 3 variáveis independentes são estatisticamente significativos a uma significância de 1% ( $p < 0,01$ ), logo existe uma relação significativa entre os descritores moleculares e a atividade biológica. A tabela 4 mostra o resultado do bom ajustamento do modelo. O Apêndice V apresenta uma análise dos resíduos do modelo obtido, que são a diferença entre o valor observado e o valor predito na amostra de treinamento. Os resultados observados nos gráficos do apêndice V mostram um comportamento que evidencia o bom ajuste do modelo aos dados do *training set*.

O gráfico do Apêndice VI, apresenta os valores observados do log IC<sub>50</sub> e os valores preditos pelo modelo para cada observação da amostra de treinamento, juntamente com um Intervalo de Predição de 95%. A equação da reta e o seu  $R^2$ , medida da qualidade do seu ajuste, são apresentados no gráfico, juntamente com o Coeficiente de Correlação de Pearson ( $R$ ) entre os valores observados e preditos, e o seu  $p$ -valor ( $p$ ). Observamos um alto  $R^2$  (0,9012) e uma correlação alta e estatisticamente significativa entre os valores observados e preditos na amostra de treinamento, de forma que o modelo prediz bem os dados de treinamento, fato também evidenciado pela reta ajustada estar próxima da reta  $y=x$ .

Por fim, temos que, para o conjunto de treinamento, o  $Q^2$  foi igual a 0.8612, o que é mais uma evidência da qualidade de predição do modelo.

### 3.2.4 Validação externa

A tabela 6 apresenta os resultados observados e calculados para a atividade biológica do *test set*. O gráfico 1 apresenta a dispersão dos resultados, sendo que a correlação de Pearson foi de  $R = 0,9643$  e o coeficiente de determinação  $R^2$  foi de 0,9298. Apesar da presença de um composto com atividade biológica muito elevada (composto 1) cuja faixa de tolerância não cruzou a reta pontilhada. A qualidade preditiva do modelo clássico se apresentou tão robusta que, o indicador de Stone-Geisser  $Q^2$  foi calculado em 0,720.

**Tabela 3** Valores observados e calculados da atividade biológica das moléculas do *training set*.

Molécula <sup>a</sup>	Log IC <sub>50</sub> Obs. <sup>b</sup>	LogIC <sub>50</sub> Calc. <sup>c</sup>	DP <sup>d</sup>
2	-0,699	-0,712	0,0092
3	-0,398	-0,513	0,0813
4	-0,004	0,413	0,2949
5	0,017	-0,049	0,0467
7	0,276	0,150	0,0891
8	0,310	0,497	0,1322
9	0,487	0,760	0,1930
10	0,713	0,813	0,0707
12	0,797	0,911	0,0806
13	0,828	0,821	0,0049
14	0,869	0,810	0,0417
15	0,889	0,867	0,0156
17	0,989	1,177	0,1329
18	1,142	1,179	0,0262
19	1,150	1,230	0,0566
20	1,189	0,921	0,1895
22	1,246	0,816	0,3041
23	1,498	1,409	0,0629
24	1,590	1,388	0,1428

<sup>a</sup>moléculas ordenadas em ordem decrescente de atividade biológica molecular (Log IC<sub>50</sub>).

<sup>b</sup>atividade biológica inibitória observada por He, Alian e Montellano [16].

<sup>c</sup> atividade biológica inibitória calculada com uso do programa BuildQSAR[23]

<sup>d</sup> desvio padrão calculado.

Fonte: Elaborado pelo autor.

**Tabela 4** parâmetros do modelo.

Parâmetro	Estimativa	Erro-padrão	Valor -t	p
Intercepto	5,426	0,428	12,689	<0,001
HA	0,173	0,048	3,629	0,002
RB	0,389	0,079	4,901	<0,001
VWallsV	-0,020	0,002	-11,567	<0,001

Fonte: elaborado pelo autor.

**Tabela 5** Valores calculados para os descritores moleculares dos compostos do *test set*.

Molécula <sup>a</sup>	HA <sup>b</sup>	RB <sup>c</sup>	VWALLS V <sup>d</sup>
1	2	2	329,56
6	5	2	318,93
11	2	2	294,98
16	2	2	283,85
21	2	2	270,31

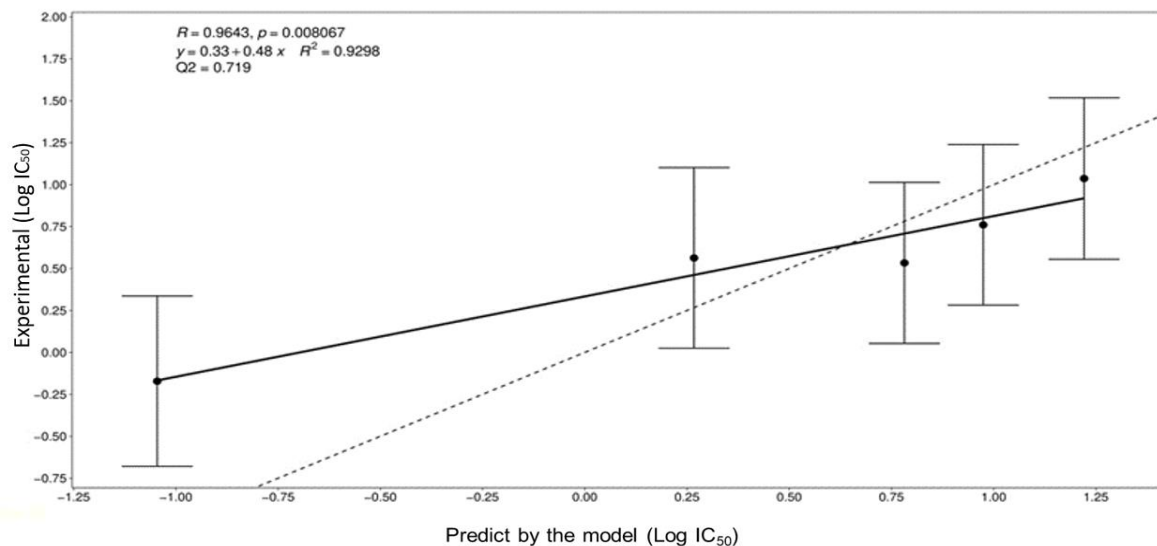
<sup>a</sup> classificação por atividade inibitória decrescente da molécula.

<sup>b</sup> número de aceptores de hidrogênio da molécula.

<sup>c</sup> número de ligações rotáveis da molécula.

<sup>d</sup> volume de Van der Walls em angstrom cúbico (Å<sup>3</sup>).

Fonte: Elaborado pelo autor.

**Gráfico 1.** Dispersão dos resultados calculados e observados da atividade biológica (Log IC<sub>50</sub>) do *test set* Clássico.

Fonte: Elaborado pelo autor

**Tabela 6** Valores observados e calculados da atividade biológica das moléculas do *test set*.

Molécula <sup>a</sup>	Log IC <sub>50</sub> Obs. <sup>b</sup>	LogIC <sub>50</sub> Calc. <sup>c</sup>	DP <sup>d</sup>
1	-1,046	-0,175	0,616
6	0,267	0,560	0,207
11	0,782	0,531	0,178
16	0,975	0,758	0,153
21	1,221	1,034	0,132

<sup>a</sup>moléculas ordenadas em ordem decrescente de atividade biológica (Log IC<sub>50</sub>).

<sup>b</sup>atividade biológica inibitória observada por He, Alian e Montellano[16].

<sup>c</sup> atividade biológica inibitória calculada com uso da função QSAR.

<sup>d</sup> desvio padrão calculado.

Fonte: Elaborado pelo autor.

### 3.2.5 Virtual Screening

A busca por compostos com o mesmo núcleo de piperazinas (Figura 2) resultou em um conjunto de 135 análogos. Foram calculados os descritores moleculares HA, RB e VVWALLS das moléculas triadas com o uso do programa MarvinView 17.24.0 [21]. Os valores calculados dos DM foram aplicados na função QSAR (Equação 1). E os 10 melhores resultados calculados foram ordenados em valores decrescentes conforme é mostrado na tabela 7.

### 3.2.6. Docagem molecular

Os 10 compostos selecionados por triagem virtual foram submetidos a docagem molecular, utilizando como ligando a enzima InhA depositada na base dados PDB sob código 1P44. Os resultados deste processo constam na tabela 8. Nota-se que a energia de ligação do composto CID 4005008 a enzima InhA foi mais baixa que a obtida para o ligante-modelo (CID 447767) do complexo depositado na base de dados PDB (1P44). O EFEB do composto triado foi de -13,3 kcal/mol, já o composto depositado na base de dados PDB teve o EFEB de -11,1 kcal/mol. O que indica uma maior afinidade teórica da molécula CID 4005008 com a InhA. As demais moléculas testadas também apresentaram baixas energias de ligação, o que apóia a

indicação de testes *in vitro* de todo o conjunto selecionado pelo modelo QSAR. Na imagem 7 podemos observar o composto CID 4005008 complexado a InhA em seu sítio catalítico. A análise das interações energéticas da molécula CID 4005008 com o EFEB mais favorável, confirmou a sua localização e interação com os resíduos de aminoácidos catalíticos, em especial Phe149 e Lys165 conforme figura 9. Também notamos que a energia de repulsão, bem como impedimentos rotacionais foram mínimos quando comparados com a energia total de afinidade como pode ser observado na figura 10, corroborando o potencial do composto CID 4005008 para eventual teste *in vitro*.

**Tabela 7** Compostos localizados por busca virtual e valores de Log IC<sub>50</sub> calculados com a função QSAR obtida neste estudo.

Molécula <sup>a</sup>	CID <sup>b</sup>	LogIC <sub>50</sub> <sup>c</sup>
1	4005008	-1,345
2	70154773	-0,771
3	10179644	-0,770
4	16764927	-0,707
5	5274981	-0,681
6	110399797	-0,570
7	78897937	-0,523
8	118725820	-0,517
9	118725821	-0,516
10	9517736	-0,456

<sup>a</sup> moléculas classificadas em ordem decrescente de atividade biológica.

<sup>b</sup>código identificador da molécula depositada na base de dados PubChem [17].

<sup>c</sup>atividade biológica medida em logaritmo de micromolar de substância necessária para atingir 50% da concentração inibitória máxima (IC<sub>50</sub>).

Fonte: Elaborado pelo autor.

**Tabela 8.** Energia livre de ligação estimada, obtida durante a docagem molecular.

Molécula (CID)	Energia de ligação <sup>a</sup>
4005008	-13,3
5274981	-12,0
70154773	-11,9
10179644	-11,7
16764927	-11,6
9517736	-11,5
118725820	-11,5
118725821	-11,0
7887937	-10,6
110399797	-9,4

<sup>a</sup>Energia de ligação estimada medidos em kcal/mol.**Tabela 9** Resultados experimentais e calculados do *training set* CoMFA.

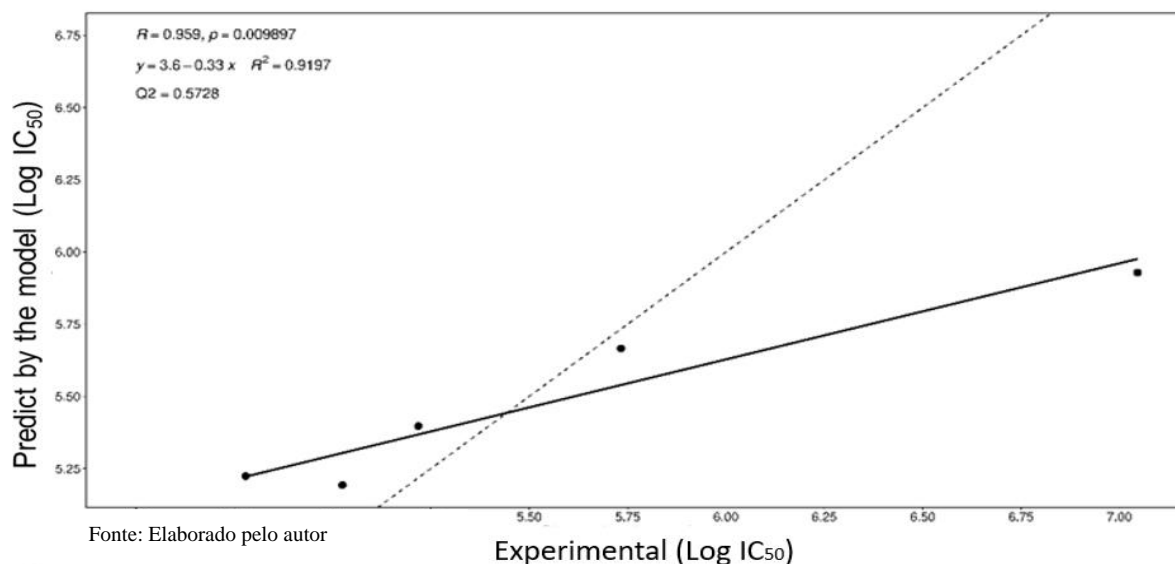
Molécula (CID) <sup>a</sup>	Exp <sup>b</sup>	Calc <sup>c</sup>
116119	4,502	4,680
258240	4,410	4,616
447767	6,699	6,720
667749	4,850	4,797
679013	4,754	5,153
681767	5,011	4,823
702506	5,172	4,825
739897	4,858	4,859
749123	5,513	5,473
788974	6,004	5,697
807901	5,203	5,172
814705	5,111	4,985
1378917	5,131	5,183
1415427	5,287	5,357
2838469	4,811	5,103
9003925	6,398	6,433
9378917	5,690	5,334
17312855	5,724	5,824
25093354	5,983	6,069

<sup>a</sup>moléculas identificadas pelo código CID (PubChem).<sup>b</sup>atividade biológica experimental convertidos pela função " $p = -\log(x)(\text{logística})$ ", a faixa desses valores logísticos abrange 2 valores de log [30].<sup>c</sup>atividade biológica calculada pelo modelo CoMFA.

### 3.3 CoMFA

#### 3.3.1 Validação interna

O *training set* foi submetido a modelagem CoMFA com as configurações citadas na metodologia obtendo

**Gráfico 2.** Dispersão dos resultados calculados e observados da atividade biológica (Log IC<sub>50</sub>) do *test set* CoMFA.

como atividade biológica estimada os valores que constam na tabela 9. O apêndice VII apresenta o gráfico do experimento e os valores preditos pelo modelo para cada observação da amostra de treinamento. Observamos um  $R^2$  e uma correlação  $R$  alta e estatisticamente significativo, a uma significância estatística de 1% ( $p < 0.01$ ), entre os valores observados e preditos por validação cruzada na amostra de treinamento. O  $Q^2$  obtido foi igual a 0,6517, o que é uma evidência da boa qualidade do modelo.

#### 3.3.2 Validação externa

Com o conjunto teste constatamos, neste caso, uma alta correlação entre os valores observados no experimento SAR, e preditos pelo modelo. Entretanto, o  $Q^2$  não é alto como no conjunto treinamento, sendo igual a 0.5728, o que representa uma qualidade preditiva apenas razoável do modelo na amostra de teste. Ainda, a reta ajustada desvia bastante da reta  $y=x$ , principalmente pelo ponto que teve valor experimental de 7,046 (Tabela 10). Os valores preditos e calculados podem ser vistos na tabela 10 e o gráfico de dispersão pode ser conferido no gráfico 2.

**Tabela 10** Resultados experimentais e calculados do *test set* CoMFA

Molécula (CID) <sup>a</sup>	Exp <sup>b</sup>	Calc <sup>c</sup>
749125	4,779	5,224
1012371	5,733	5,666
1070143	5,218	5,397
4006554	5,025	5,193
8366407	7,046	5,929

<sup>a</sup>moléculas identificadas pelo código CID (PubChem).<sup>b</sup>atividade biológica experimental convertidos pela função " $p = -\log(x)(\text{logística})$ ", a faixa desses valores logístico abrange 2 valores de log.<sup>c</sup>atividade biológica calculada pelo modelo CoMFA.

### 3.3.4 Campos de força utilizados

Sugere-se que com o conjunto de moléculas do SAR mais as configurações da modelagem utilizadas, a energia de interação baseada na estereoquímica foi a mais relevante para correlacionar-se com a atividade biológica do que a energia eletrostática, ou ambas energias somadas. Esta conclusão pode ser entendida ao observar o quadro 1, onde podemos avaliar a qualidade preditiva do modelo CoMFA utilizando somente valores de energia de interação estereoquímica (STE), somente valores de energia de interação eletrostática (ELE) ou ambas energias (STE-ELE). O quadro 1 nos levou a considerar somente a energia STE para modelar o *training set* e o *test set* sendo que, todos os resultados CoMFA citados nesta pesquisa foram obtidos utilizando os dados da energia de interação STE.

Quadro 1

	STE <sup>a</sup>	ELE <sup>b</sup>	STE-ELE <sup>c</sup>
$Q^2$ <sup>d</sup>	0,6517	0,0720	0,3250

<sup>a</sup> energia estereoquímica.

<sup>b</sup> energia eletrostática.

<sup>c</sup> ambas energias.

<sup>d</sup> indicador de Stone-Geisser  $Q^2$

### 3.3.5 Geração de campos STE 3D

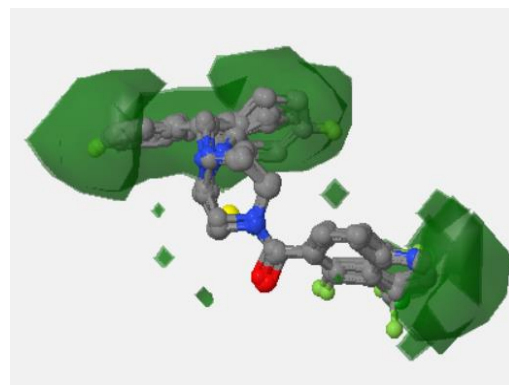
Com a modelagem CoMFA concluída, foram gerados os gráficos tridimensionais indicando a localização onde a energia de interação STE é mais favorável e menos favorável. Esta é uma informação relevante para o aperfeiçoamento molecular. Na figura 6 podemos identificar onde as interações STE são mais favoráveis, já na figura 7 observamos onde as interações STE são menos favoráveis. De acordo com Ankur Vaidya *et al* [31] campos energéticos STE volumosos (figura 6) favorecem a atividade biológica, logo o estudo de uma substituição hidrofóbica volumosa nestas regiões teoricamente poderia aumentar a atividade inibitória desejada.

### 3.3.6 Discussão

Notamos que o modelo QSAR clássico obteve uma análise estatística mais favorável que o modelo CoMFA. Esse resultado talvez esteja relacionado ao fato que a modelagem clássica, usando RLM, conseguiu reproduzir melhor um resultado extremo, referente a molécula 1 do conjunto teste, que o CoMFA utilizando PLS. Contudo essa pesquisa decidiu por não executar um indesejado tratamento estatístico de *outliers*, como uma forma de avaliar com mais clareza os 2 modelos em questão. A discussão de retirar ou não *outliers* é válida, pois tanto o CoMFA quanto o QSAR clássico obteriam resultados excepcionais sem o ponto “extremo”, entretanto é desejável buscar um modelo

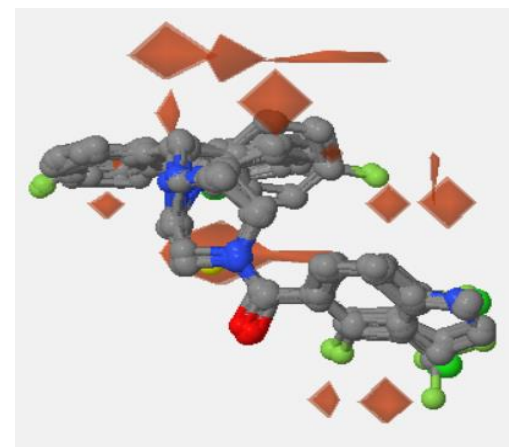
capaz de reproduzir ou prever razoavelmente os dados da realidade, mesmo com aqueles valores que fogem da média. Apesar de um *outlier* eventualmente criar um viés negativo em uma análise, pode ser esse mesmo *outlier* a resposta para o problema em estudo

Figura 6. Mapa de contorno CoMFA STE favorável



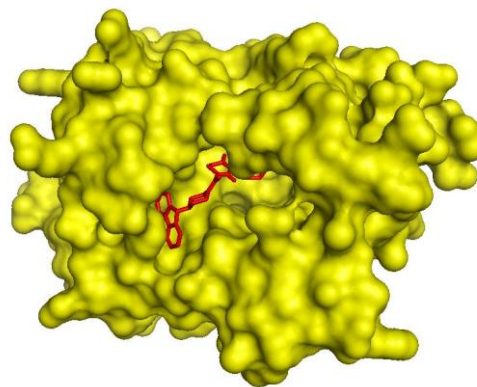
Interações estereoquímicas favoráveis.

Figura 7. Mapa de contorno CoMFA STE desfavorável



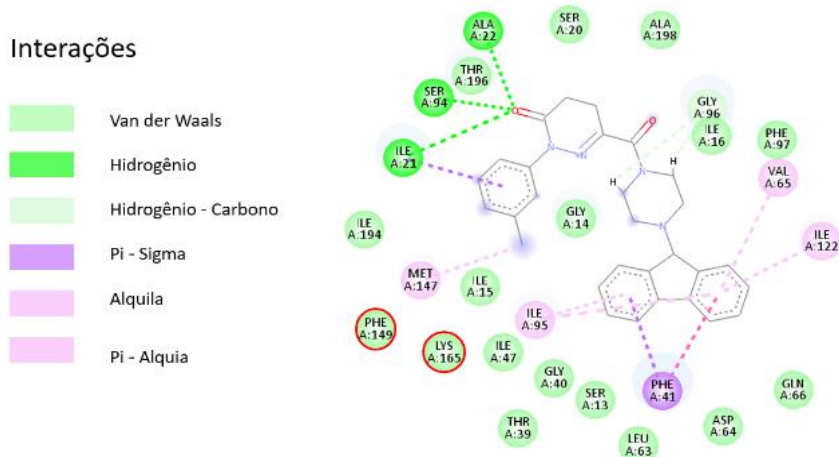
Interações estereoquímicas desfavoráveis.

Figura 8. Composto CID 4005008 (em vermelho) complexado a InhA (em amarelo).



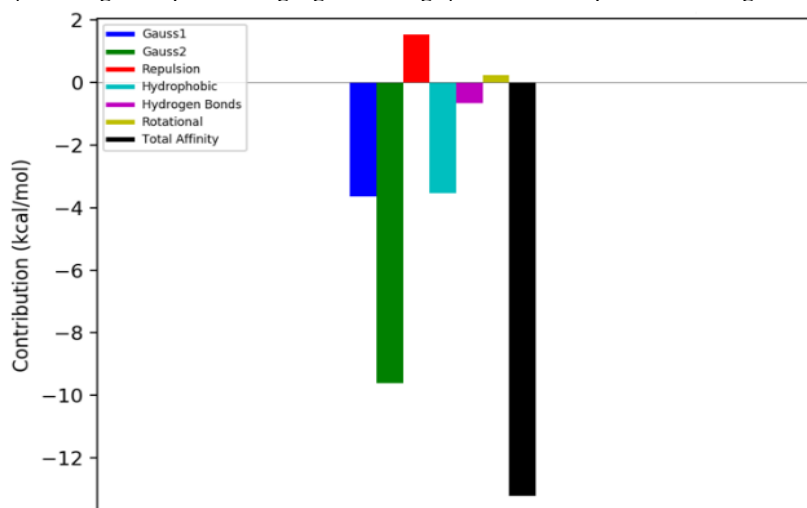
Fonte: Imagem do autor.

**Figura 9.** Interações energéticas da molécula CID 4005008 no sítio catalítico da InhA. Destacando em círculo vermelho, os resíduos de aminoácidos PHE149 e LYS 165, importantes para a atividade biológica.



Fonte: Elaborado pelo autor com BIOVIA [28].

**Fig. 10** Contribuições energéticas para a energia global de ligação entre o receptor InhA e o ligante CID 4005008.



Fonte: obtido em Achilles blind docking server [29].

#### 4 Conclusão

Nesta pesquisa, observamos a capacidade da modelagem QSAR clássica de gerar uma função que estime a atividade inibitória de uma série de análogos de arilamidas, em termos de seus descritores moleculares. Foi possível obter uma função que, tanto na validação interna quanto na externa, geraram ótimos resultados. Tais resultados permitiram o uso da Função QSAR clássico para efetuar uma triagem de moléculas estruturalmente semelhantes as utilizadas por He e colaboradores [16]. Das 135 moléculas obtidas em triagem virtual e submetidas a função QSAR, utilizamos os 10 compostos com a maior atividade biológica calculada, para uma etapa de docagem molecular que, não apenas apontou a modelagem QSAR clássico apropriada para localizar moléculas com baixa energia de ligação estimada,

como também encontrou compostos teoricamente com mais afinidade a enzima InhA do que o complexo ligante/ligado originalmente depositado na base de dados PDB usado como modelo.

A aplicação do modelo matemático CoMFA obteve bons resultados utilizando os dados das interações energéticas STE, permitindo gerar o gráfico tridimensional do campo de interação da energia esteroquímica no entorno das moléculas do estudo de SAR. Essas informações possibilitaram propor substituições hidrofóbicas volumosas que teoricamente poderiam aumentar a atividade biológica dos compostos *drug-like*. A utilização dos 2 modelos QSAR propiciou ter um quadro mais amplo dos fatores que favorecem ou desfavorecem a atividade biológica no sítio catalítico da InhA, além de sugerir novas moléculas candidatas a ensaios *in vitro*. A pesquisa

constatou que usar os 2 modelos matemáticos nos oferece respostas mais abrangentes para a busca de novos compostos de interesse, o que pode ser uma proposta para não limitar futuras pesquisas nessa área a somente 1 modelo preditivo.

#### DECLARAÇÃO DE CONFLITO DE INTERESSES

Os autores declaram que não possuem conflito de interesses.

#### REFERÊNCIAS

- [1] World Health Organization (WHO). Global tuberculosis report 2020: executive summary. Geneva; 2020. Disponível em: <<https://www.who.int/publications-detail-redirect/978924001313>>.
- [2] Pedrique B, Strub-Wourgaft N, Some C, Olliaro P, Trouiller P, Ford N, et al. The drug and vaccine landscape for neglected diseases (2000–11): a systematic assessment. *The Lancet Global Health*. 2013 Dec;1(6):e371–9. Disponível em: <[http://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(13\)70078-0/abstract](http://www.thelancet.com/journals/langlo/article/PIIS2214-109X(13)70078-0/abstract)>. Acesso em: 10 de maio de 2021. doi: 10.1016/S2214-109X(13)70078-0
- [3] Brigden G, Castro J, Ditiu L, Gray G, Hanna D, Low M, et al. Tuberculosis and antimicrobial resistance – new models of research and development needed. *Bulletin of the World Health Organization* [Internet]. 2017 May 1;95(5):315–5. doi:10.2471/blt.17.194837.
- [4] Golan DE. Princípios de farmacologia: a base fisiopatológica da farmacoterapia. Rio De Janeiro: Guanabara Koogan; 2009.
- [5] Ferreira MMC, Montanari CA, Gaudio AC. Seleção de variáveis em QSAR. *Química Nova*. 2002 May;25(3):439–48.
- [6] Narasimhan B. Qsar by hansch analysis. *J Pharm BioSci*. 2014;2(4):2-6
- [7] Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*. 1988 Aug;110(18):5959–67. doi: 10.1021/ja00226a005.
- [8] Nelson DL, Cox MM. Princípios de bioquímica de Lehninger. 5th ed. Porto Alegre: Artmed; 2011.
- [9] Chollet A, Maveyraud L, Lherbet C, Bernardes-Génisson V. An overview on crystal structures of InhA protein: Apo-form, in complex with its natural ligands and inhibitors. *European Journal of Medicinal Chemistry*. 2018 Feb;146:318–43.
- [10] Morlock GP, Metchock B, Sikes D, Crawford JT, Cooksey RC. ethA , inhA , and katG Loci of Ethionamide-Resistant Clinical Mycobacterium tuberculosis Isolates. *Antimicrobial Agents and Chemotherapy*. 2003 Dec;47(12):3799–805. doi: 10.1128/AAC.47.12.3799-3805.2003.
- [11] Freundlich Joel S, Wang F, Vilchère C, Gulten G, Langley R, Schiehser Guy A, et al. Triclosan Derivatives: Towards Potent Inhibitors of Drug-Sensitive and Drug-Resistant Mycobacterium tuberculosis. *ChemMedChem*. 2009 Feb 13;4(2):241–8. doi: 10.1002/cmdc.200800261
- [12] Khan R, Zeb A, Roy N, Magar RT, Kim HJ, Lee KW, et al. Biochemical and Structural Basis of Triclosan Resistance in a Novel Enoyl-Acyl Carrier Protein Reductase. *Antimicrobial Agents and Chemotherapy* [Internet]. 2018 Aug 1;62(8). doi:10.1128/AAC.00648-18
- [13] Khade AB, Boshoff HIM, Arora K, Vandana KE, Verma R, Shenoy GG. Design, synthesis, evaluation, and molecular dynamic simulation of triclosan mimic diphenyl ether derivatives as antitubercular and antibacterial agents. *Structural Chemistry*. 2020 Jan 6;31(3):983–98. doi:10.1007/s11224-019-0147
- [14] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. 2019 Jan 8;47(Database issue): D607–13.
- [15] Vilchère C, Jacobs JR. WR. Resistance to Isoniazid and Ethionamide in Mycobacterium tuberculosis: Genes, Mutations, and Causalities. *Microbiology Spectrum*. 2014 Aug 22;2(4).
- [16] He X, Alian A, Ortiz de Montellano PR. Inhibition of the Mycobacterium tuberculosis enoyl acyl carrier protein reductaseInhA by arylamides. *Bioorganic & Medicinal Chemistry*. 2007 Nov;15(21):6649–58. doi:10.1016/j.bmc.2007.08.013
- [17] Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. *Nucleic Acids Research* [Internet]. 2015 Sep 22;44(D1): D1202–13.
- [18] Martin TM, Harten P, Young DM, Muratov EN, Golbraikh A, Zhu H, et al. Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? *Journal of Chemical Information and Modeling*. 2012 Oct 3;52(10):2570–8. doi: 10.1021/ci300338w
- [19] Leonard JT, Roy K. On Selection of Training and Test Sets for the Development of Predictive QSAR models. *QSAR & Combinatorial Science*. 2006 Mar;25(3):235–51. doi: 10.1002/qsar.200510161
- [20] Damre MV, Gangwal RP, Dhoke GV, Lalit M, Sharma D, Khandelwal K, et al. 3D-QSAR and molecular docking studies of amino-pyrimidine derivatives as PknB inhibitors. *Journal of the Taiwan Institute of Chemical Engineers*. 2014 Mar;45(2):354–64. doi: 10.1016/j.jtice.2013.05.016

- [21] Chemaxon LTD., 2017. MarvinView 17.24.0 para Windows. Hungría: ChemAxon. Conjunto de datos. Ambiente operacional.
- [22] Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, et al. Virtual Computational Chemistry Laboratory – Design and Description. *Journal of Computer-Aided Molecular Design*. 2005 Jun;19(6):453–63.
- [23] Oliveira DB, Gaudio AC. BuildQSAR: A New Computer Program for QSAR Analysis. *Quantitative Structure-Activity Relationships*. 2000 Dec;19(6):599–601. doi:10.1002/1521-3838(200012)19:6<599::AID-QSAR599>3.0.CO;2-B
- [24] Neter J, Kutner M, Wasserman W, Nachtsheim C. *Applied linear statistical models*. 4th ed. Boston: McGraw-Hill/Irwin, 1996. 1415p.
- [25] Chin WW. How to Write Up and Report PLS Analyses. *Handbook of Partial Least Squares* [Internet]. 2009 Nov 16;655–90. doi: 10.1007/978-3-540-32827-8\_29
- [26] Berman HM. The Protein Data Bank. *Nucleic Acids Research*. 2000 Jan 1;28(1):235–42. doi: 10.1093/nar/28.1.235
- [27] Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol*. 2015; 1263:243-250. doi:10.1007/978-1-4939-2269-7\_19
- [28] Biovia. Dassault Systèmes. *Discovery Studio Visualizer*, V21.1.0.20298, San Diego: 2021.
- [29] Sánchez-Linares I, Pérez-Sánchez H, Cecilia JM, García JM. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinformatics*. 2012;13(Suppl 14): S13. doi: 10.1186/1471-2105-13-S14-S13
- [30] Ragno R. www.3d-qsar.com: a web portal that brings 3-D QSAR to all electronic devices—the Py-CoMFA web application as tool to build models from pre-aligned datasets. *Journal of Computer-Aided Molecular Design*. 2019 Sep;33(9):855–64.
- [31] Vaidya A, Jain AK, Prashantha Kumar BR, Sastry GN, Kashaw SK, Agrawal RK. CoMFA, CoMSIA, kNN MFA and docking studies of 1,2,4-oxadiazole derivatives as potent caspase-3 activators. *Arabian Journal of Chemistry*. 2017 May;10: S3936–46.
- [32] Todeschini R, Consonni V. *Handbook of molecular descriptors*. Weinheim; New York: Wiley-Vch; 2000.

## Apêndice I

Tabela 11 Descritores moleculares avaliados neste estudo.

Descritor molecular	Descrição*
<b>MW</b>	Massa molecular média calculada a partir dos pesos atômicos padrão.
<b>MW Exact</b>	Massa monoisotópica calculada a partir dos pesos dos isótopos naturais mais abundantes dos elementos.
<b>Atomcount</b>	Número de átomos na molécula
<b>Complexity</b>	Leva em consideração tanto a variedade de tipos de conectividade quanto tipos de átomos de um gráfico molecular sem H +.
<b>HA</b>	Aceptores de hidrogênio
<b>RB</b>	Ligações rotáveis na molécula
<b>XLogP3</b>	Classificação otimizada com esquema de 87 tipos de átomos / grupos, bem como dois fatores responsáveis por ligações H internas e aminoácidos
<b>Log P</b>	Coefficiente de partição octanol / água.
<b>HLB</b>	Balanco hidrófilo-lipofílico, grau hidrofílico ou lipofílico da molécula.
<b>MPLZ</b>	Polarizabilidade molecular, tendência relativa de uma nuvem de elétrons (uma distribuição de carga) de uma molécula a ser distorcida por um campo elétrico externo.
<b>Weinerpolarizability</b>	O número de 3 corpos de distância da molécula.
<b>Dreidingenergy</b>	Energia da estrutura 3D (conformação) da molécula usando o campo de força Dreiding.
<b>MMFF94</b>	Energia da estrutura 3D (conformação) da molécula usando o campo de força MMFF94.
<b>Van der Walls</b>	Volume Van der Waals do <i>conformer</i> .
<b>Volume</b>	
<b>PSA</b>	A área de superfície polar (PSA) é formada por átomos polarizados da molécula.
<b>MSA 3D</b>	Área de superfície molecular 3D.
<b>RF</b>	Refratividade, relacionado ao volume das moléculas e às forças dispersivas de London
<b>Log S</b>	Solubilidade em água.

Fonte: Programa MarvinView 17.24.0 [21,32] e Todeschini e Consonni [32]

## Apêndice II

Tabela 12 Série de arilamidas testadas *in vitro* por He et al [16].

Molécula*	CID **	IC <sub>50</sub> µm***
1	8366407	0,09
2	447767	0,2
3	9003925	0,4
4	788974	0,99
5	25093354	1,04
6	1012371	1,85
7	17312855	1,89
8	9378917	2,04
9	749123	3,07
10	1415427	5,16
11	1070143	6,05
12	807901	6,26
13	702506	6,73
14	1378917	7,39
15	814705	7,74
16	4006554	9,43
17	681767	9,74
18	739897	13,87
19	667749	14,11
20	2838469	15,47
21	749125	16,64
22	679013	17,62
23	116119	31,5
24	258240	38,86

\*moléculas classificadas em ordem decrescente de atividade inibitória. \*\*código identificador da molécula depositado na base de dados PubChem [17]. \*\*\*atividade biológica medida em micromolar de substancia necessária para atingir 50% da concentração inibitória máxima (IC<sub>50</sub>).

Fonte: He; Alian; Montellano, 2007 [16].

## Apêndice III

Tabela 13 Composição dos conjuntos *Test set* e *Training set*.

<i>Data set</i>		
<b>Molécula*</b>	<i>Test set</i> (CID)	<i>Training set</i> (CID)
1	8366407	
2		447767
3		9003925
4		788974
5		25093354
6	1012371	
7		17312855
8		9378917
9		749123
10		1415427
11	1070143	
12		807901
13		702506
14		1378917
15		814705
16	4006554	
17		681767
18		739897
19		667749
20		2838469
21	749125	
22		679013
23		116119
24		258240

\* Moléculas que compõem o estudo de SAR, ordenadas por ordem decrescente de atividade biológica. Os compostos estão divididos em colunas *test set* e *training set*, e identificadas pelo código CID utilizado na base dados PubChem. Fonte: Elaborado pelo autor.

## Apêndice IV

Tabela 14 Valores calculados para os descritores moleculares do conjunto treino.

Molécula <sup>a</sup>	HA <sup>b</sup>	RB <sup>c</sup>	VWALLS V <sup>d</sup>
2	2	2	356,067
3	2	2	346,311
4	2	2	300,899
5	4	4	378,613
7	4	4	368,857
8	4	4	351,813
9	2	2	283,854
10	1	3	291,845
12	5	2	301,864
13	2	2	280,899
14	1	3	292,032
15	1	3	289,196
17	3	2	271,895
18	3	2	271,782
19	2	3	279,875
20	4	2	292,928
22	2	2	281,106
23	2	3	271,115
24	2	2	253,055

<sup>a</sup> classificação por atividade molecular inibitória decrescente.

<sup>b</sup> número de aceptores de hidrogênio da molécula.

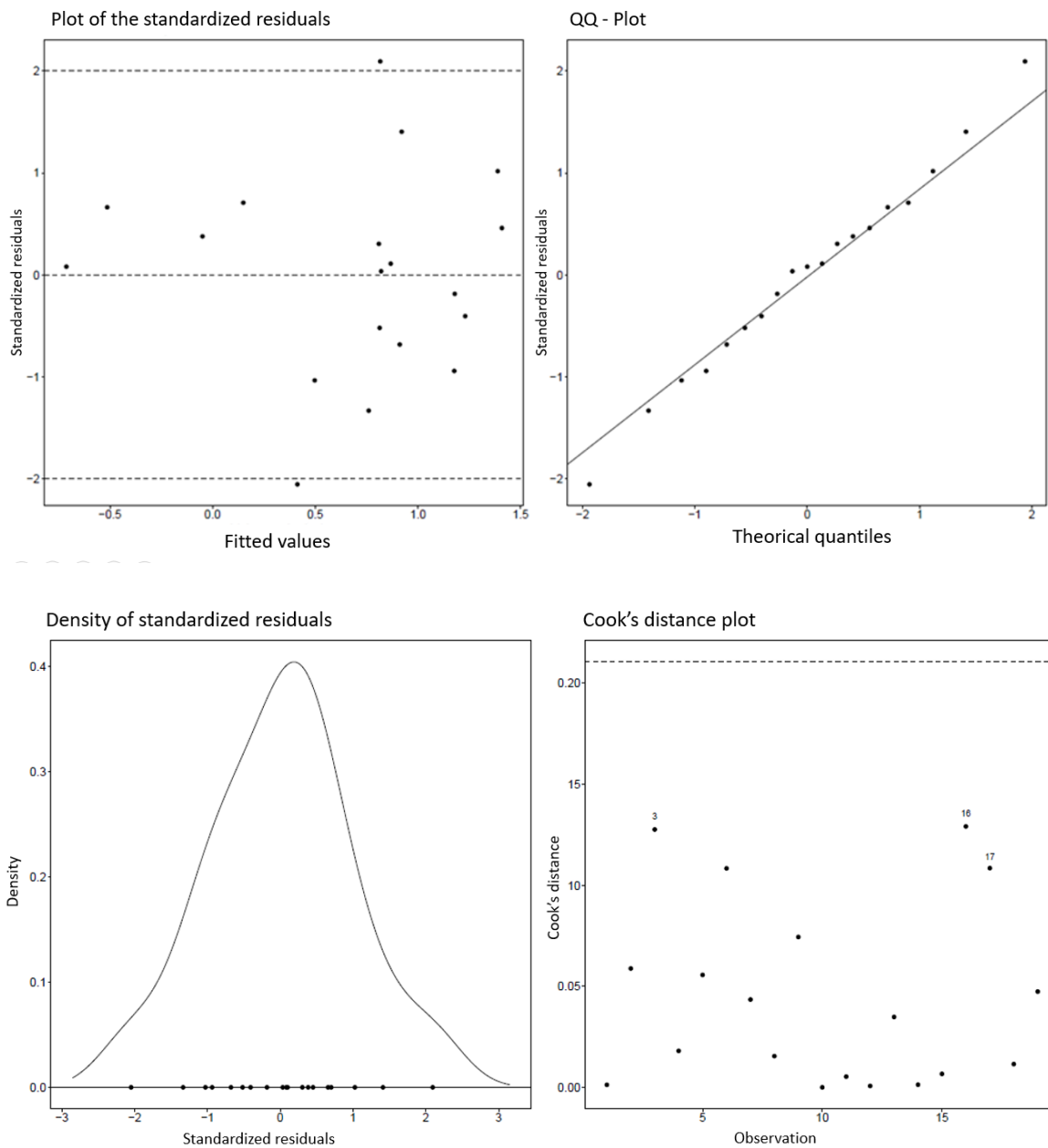
<sup>c</sup> número de ligações rotáveis da molécula.

<sup>d</sup> Volume de Van der Walls em angstrom cúbico (Å<sup>3</sup>).

Fonte: Elaborado pelo autor.

## Apêndice V

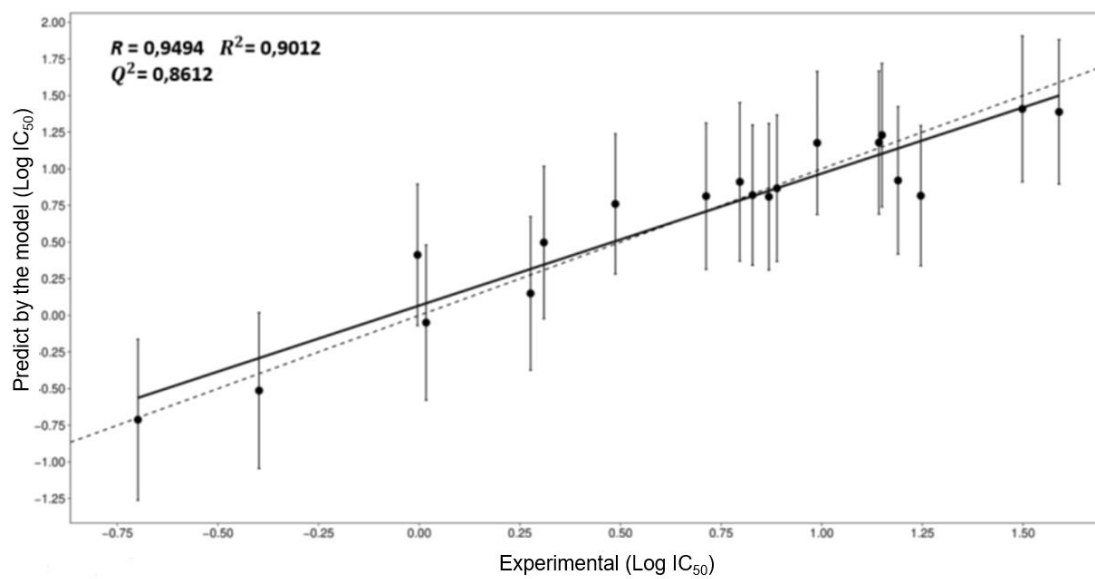
Fig. 11 Análise de resíduos do modelo ajustado.



Fonte: Elaborado pelo autor.

## Apêndice VI

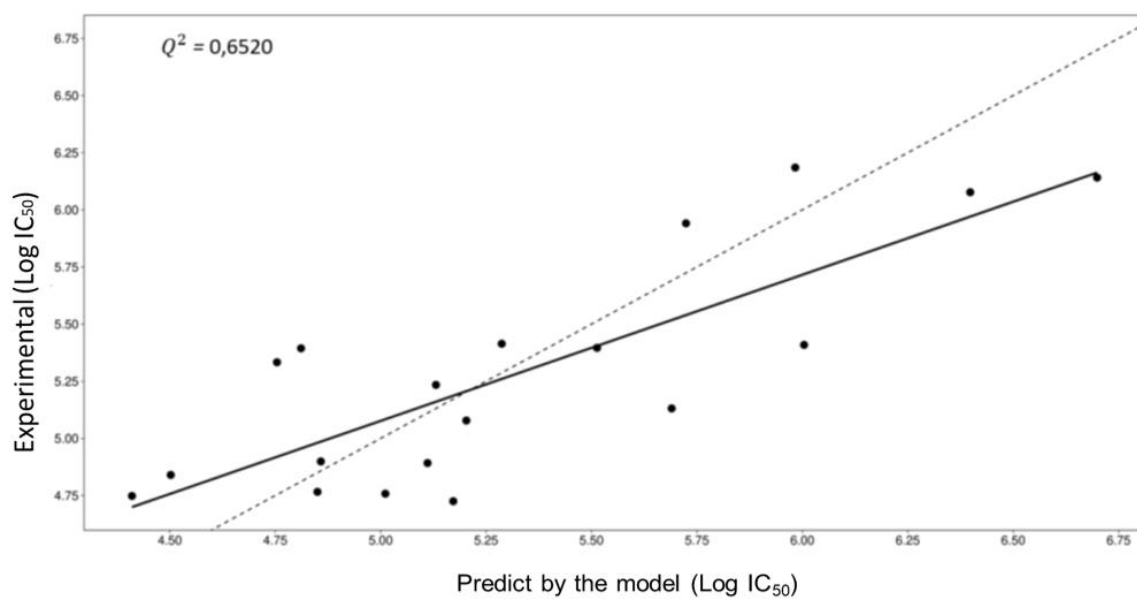
**Gráfico 4.** Dispersão dos resultados calculados e observados da atividade biológica (Log IC<sub>50</sub>) do *training set* na modelagem QSAR clássica.



Fonte: Elaborado pelo autor.

## Apêndice VII

**Gráfico 5** Dispersão dos resultados calculados e observados da atividade biológica (Log IC<sub>50</sub>) do *training set* na modelagem QSAR CoMFA.



Fonte: elaborado pelo autor.

## 5 DISCUSSÕES

Notamos que o modelo QSAR clássico, por fornecer a função QSAR, permitiu uma análise estatística mais completa que o modelo CoMFA. A ferramenta CoMFA gerou muitas informações estatísticas importantes nos seus resultados, mas limita uma análise externa de adequação do modelo devido à falta da função QSAR gerada. Tanto a modelagem clássica usando RLM quanto o CoMFA utilizando PLS, apresentam alguma dificuldade de reproduzir ou prever valores extremos, no entanto essa pesquisa decidiu por não executar um indesejado tratamento estatístico de *outliers* como uma forma de avaliar com mais clareza os 2 modelos em questão. A discussão de retirar ou não *outliers* é válida, pois tanto o CoMFA quanto o QSAR clássico obteriam resultados excepcionais sem o chamado ponto “fora da curva”. Entretanto é desejável ter um modelo capaz de reproduzir ou prever os dados da realidade, mesmo com pontos que fogem da média. Apesar de um *outlier* eventualmente criar um viés negativo em uma análise, pode ser esse mesmo *outlier* a resposta para o problema em estudo. Neste caso concluímos que os 2 modelos obtiveram resultados preditivos relevantes, mesmo incluindo um ponto extremo. Os resultados destacaram o modelo Clássico com bom ajuste de dados e ótima capacidade preditiva. Já o modelo CoMFA forneceu dados importantes para um aperfeiçoamento molecular, usado para otimizar as moléculas candidatas a testes *in vitro*. Usar os 2 modelos matemáticos preditivos nos deu uma resposta mais ampla para a busca de novos compostos *drug-like*, o que pode ser uma indicação para não limitar futuras pesquisas a somente 1 modelo QSAR.

## 6 CONCLUSÕES

Esta pesquisa comprovou a viabilidade do desenvolvimento de modelos QSAR para estimar a atividade biológica inibitória de moléculas, tendo como alvo a enzima InhA. O modelo Hansh-Fujita alcançou elevada correlação com resultados *in vitro*, assim como elevado fator  $Q^2$ , justificando desta forma, a sua utilização como ferramenta preditiva.

A função QSAR, gerada na modelagem matemática, avaliou 135 moléculas análogas a molécula melhor ranqueada do conjunto treino, esta ação possibilitou a seleção dos dez compostos mais promissores. Estes análogos, ao serem submetidos a docagem molecular, confirmaram uma baixa energia de ligação estimada entre os compostos e a enzima InhA. Já o modelo CoMFA delimitou os campos energéticos estereoquímicos, viabilizando propor aperfeiçoamento molecular sobre os análogos selecionados pelo modelo Hansh-Fujita. A molécula CID 4005008, que obteve a melhor atividade biológica inibitória estimada calculada pelo modelo QSAR, também alcançou, durante a docagem molecular, uma energia de ligação estimada menor do que a molécula melhor ranqueada do conjunto treino, o que a qualifica para uma fase de teste *in vitro*.

Os dois modelos matemáticos utilizados, comprovaram a força da metodologia QSAR para apoiar a pesquisa de novas moléculas *drug-like*.

## REFERÊNCIAS

- Amador JP, Lopes SJ, Pereira JE, Souza AM, Toebe M. Análise das Pressuposições e Adequação dos Resíduos em Modelo de Regressão Linear para Individuais, Ponderados e não Ponderados, utilizando Procedimentos do SAS®. *Ciência e Natura*. 2011, 33(2). Disponível em: <<https://periodicos.ufsm.br/cienciaenatura/article/view/9359>>.
- Andrade CH, Pasqualoto KFM, Zaim MH, Ferreira EI. Abordagem racional no planejamento de novos tuberculostáticos: inibidores da InhA, enoil-ACP redutase do *M. tuberculosis*. *Revista Brasileira de Ciências Farmacêuticas*. 2008 Jun;44(2):167–79.
- Baker L, Brown T, Maiden MC, Drobniowski F. Silent Nucleotide Polymorphisms and a Phylogeny for *Mycobacterium tuberculosis*. *Emerging Infectious Diseases*. 2004 Sep;10(9):1568–77.
- Berman HM. The Protein Data Bank. *Nucleic Acids Research*. 2000 Jan 1;28(1):235–42. doi: 10.1093/nar/28.1.235
- Brown L, Wolf JM, Prados-Rosales R, Casadevall A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nature Reviews Microbiology* [Internet]. 2015 Sep 1 [cited 2020 Jan 17];13(10):620–30. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4860279/>>.
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, et al. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology*. 2006 Mar 6;6(1). doi: 10.1186/1471-2180-6-23
- Chin WW. How to Write Up and Report PLS Analyses. *Handbook of Partial Least Squares* [Internet]. 2009 Nov 16;655–90. doi: 10.1007/978-3-540-32827-8\_29
- Chollet A, Maveyraud L, Lherbet C, Bernardes-Génisson V. An overview on crystal structures of InhA protein: Apo-form, in complex with its natural ligands and inhibitors. *European Journal of Medicinal Chemistry*. 2018 Feb;146:318–43.
- Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*. 1988 Aug;110(18):5959–67. DOI: 10.1021/ja00226a005
- Cruz DS. Estudos de QSAR 2D e 3D para derivados de aminoimidazóis, aminohidantoínas e aminopiridinas com atividade inibitória sobre a enzima beta-secretase humana [Internet] [Dissertação]. [Universidade Estadual de Feira de Santana]; 2012. p. 108. Disponível em: <<http://tede2.uefs.br:8080/handle/tede/1009>>

Duan X, Xiang X, Xie J. Crucial components of mycobacterium type II fatty acid biosynthesis (Fas-II) and their inhibitors. *FEMS Microbiology Letters*. 2014 Oct 21;360(2):87–99.

Dulberger CL, Rubin EJ, Boutte CC. The mycobacterial cell envelope — a moving target. *Nature Reviews Microbiology*. 2019 Nov 14;18(1):47–59. doi: 10.1038/s41579-019-0273-7.

Fährrolfes R, Bietz S, Flachsenberg F, Meyder A, Nittinger E, Otto T, et al. ProteinsPlus: a web portal for structure analysis of macromolecules. *Nucleic Acids Research*. 2017 May 2;45(W1):W337–43.

Ferreira MMC, Montanari CA, Gaudio AC. Seleção de variáveis em QSAR. *Química Nova*. 2002 May;25(3):439–48.

Freundlich JS, Wang F, Vilchère C, Gulten G, Langley R, Schiehser Guy A, et al. Triclosan Derivatives: Towards Potent Inhibitors of Drug-Sensitive and Drug-Resistant *Mycobacterium tuberculosis*. *ChemMedChem*. 2009 Feb 13;4(2):241–8. doi: 10.1002/cmdc.200800261

He X, Alian A, Ortiz de Montellano PR. Inhibition of the *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase InhA by arylamides. *Bioorganic & Medicinal Chemistry*. 2007 Nov;15(21):6649–58. doi:10.1016/j.bmc.2007.08.013

Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High Functional Diversity in *Mycobacterium tuberculosis* Driven by Genetic Drift and Human Demography. Blaser MJ, editor. *PLoS Biology*. 2008 Dec 16;6(12): e311. doi: 10.1371/journal.pbio.0060311

Hoffmann R. Análise de regressão: uma introdução à econometria [Internet]. 5th ed. Piracicaba; 2016. Disponível em: <[https://www.esalq.usp.br/biblioteca/sites/default/files/Analise\\_Regress%C3%A3o.pdf](https://www.esalq.usp.br/biblioteca/sites/default/files/Analise_Regress%C3%A3o.pdf)>.

Khade AB, Boshoff HIM, Arora K, Vandana KE, Verma R, Shenoy GG. Design, synthesis, evaluation, and molecular dynamic simulation of triclosan mimic diphenyl ether derivatives as antitubercular and antibacterial agents. *Structural Chemistry*. 2020 Jan 6;31(3):983–98. doi:10.1007/s11224-019-0147

Khan R, Zeb A, Roy N, Magar RT, Kim HJ, Lee KW, et al. Biochemical and Structural Basis of Triclosan Resistance in a Novel Enoyl-Acyl Carrier Protein Reductase. *Antimicrobial Agents and Chemotherapy* [Internet]. 2018 Aug 1;62(8). Disponível em: <<https://aac.asm.org/content/62/8/e00648-18>>. doi:10.1128/AAC.00648-18

Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. *Nucleic Acids Research* [Internet]. 2015 Sep 22;44(D1): D1202–13

Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied linear statistical models. 5<sup>th</sup> ed. New Delhi: Mcgraw-Hill Education (India) Private Limited; 2013.

Leonard JT, Roy K. On Selection of Training and Test Sets for the Development of Predictive QSAR models. QSAR & Combinatorial Science. 2006 Mar;25(3):235–51. doi:10.1002/qsar.200510161

Lopes IL, Menezes SA de, Dutra YS, Vandesmet LCS. Aspectos microbiológicos e imunológicos da infecção por *Mycobacterium tuberculosis*. Revista Expressão Católica Saúde. 2016 Jun 30;1(1). Disponível em: <<http://publicacoesacademicas.unicatolicaquixada.edu.br/index.php/recsaude/article/view/1366>>. Acesso em: 21 Jun. 2021. doi:10.25191/recs.v1i1.1366.

Lopes LGF, Carvalho EM, Sousa EHS. A bioinorganic chemistry perspective on the roles of metals as drugs and targets against *Mycobacterium tuberculosis* – a journey of opportunities. Dalton Transactions [Internet]. 2020 Nov 25;49(45):15988–6003. Disponível em: <<https://pubs.rsc.org/en/content/articlelanding/2020/dt/d0dt01365j#>DOI: 10.1039/d0dt01365j>

Morlock GP, Metchock B, Sikes D, Crawford JT, Cooksey RC. ethA, inhA, and katG Loci of Ethionamide-Resistant Clinical *Mycobacterium tuberculosis* Isolates. Antimicrobial Agents and Chemotherapy. 2003 Dec;47(12):3799–805.doi: 10.1128/AAC.47.12.3799-3805.2003.

MOTA, SP. Isolamento e identificação molecular de micobactérias não tuberculosas [Internet] [Dissertação]. [Universidade de Aveiro]; 2011. p. 64. Disponível em: <<https://ria.ua.pt/bitstream/10773/7736/1/5819.pdf>>.

Murray PR, Rosenthal KS, Pfaller MA, Tenover FC, Tenover FC. Microbiologia medica. Rio De Janeiro: Elsevier; 2009.

Narasimhan B. Qsar by hansch analysis. J Pharm BioSci. 2014;2(4):2-6

Nelson DL, Cox MM. Princípios de bioquímica de Lehninger. 5th ed. Porto Alegre: Artmed; 2011.

Neter J, Kutner M, Wasserman W, NachtsheimC. Applied linear statistical models. 4th ed. Boston: McGraw-Hill/Irwin, 1996. 1415p.

Nguyen QH, Contamin L, Nguyen TVA, Bañuls AL. Insights into the processes that drive the evolution of drug resistance in *Mycobacterium tuberculosis*. Evol Appl. 2018;11(9):1498-1511. Published 2018 Jun 21. doi:10.1111/eva.12654

Oliveira DB, Gaudio AC. BuildQSAR: A New Computer Program for QSAR Analysis. Quantitative Structure-Activity Relationships. 2000 Dec;19(6):599–601.doi:10.1002/1521-3838(200012)19:6<599:AID-QSAR599>3.0.CO;2-B

Pawełczyk J, Kremer L. The Molecular Genetics of Mycolic Acid Biosynthesis. *Microbiology Spectrum*. 2014 Aug 22;2(4). doi: 10.1128/microbiolspec.MGM2-0003-2013.

Pedrique B, Strub-Wourgaft N, Some C, Olliaro P, Trouiller P, Ford N, et al. The drug and vaccine landscape for neglected diseases (2000–11): a systematic assessment. *The Lancet Global Health*. 2013 Dec;1(6):e371–9. Disponível em: <[http://www.thelancet.com/journals/langlo/article/PIIS2214-109X\(13\)70078-0/Abstract](http://www.thelancet.com/journals/langlo/article/PIIS2214-109X(13)70078-0/Abstract)>. Acesso em: 10 de maio de 2021. doi: 10.1016/S2214-109X(13)70078-0

Sánchez-Linares I, Pérez-Sánchez H, Cecilia JM, García JM. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinformatics*. 2012;13(Suppl 14):S13. doi: 10.1186/1471-2105-13-S14-S13

Ragno R. [www.3d-qsar.com](http://www.3d-qsar.com): a web portal that brings 3-D QSAR to all electronic devices—the Py-CoMFA web application as tool to build models from pre-aligned datasets. *Journal of Computer-Aided Molecular Design*. 2019 Sep;33(9):855–64.

Ramakrishnan L. Revisiting the role of the granuloma in tuberculosis. *Nature Reviews Immunology*. 2012 Apr 20;12(5):352–66.

Raman, K., Chandra, N. *Mycobacterium tuberculosis* interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol* **8**, 234 (2008). <https://doi.org/10.1186/1471-2180-8-234>

Roy K, Kar S, Das RN. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*. Cham: Springer International Publishing; 2015.

Schiebel J, Kapilashrami K, Fekete A, Bommineni GR, Schaefer CM, Mueller MJ, et al. Structural Basis for the Recognition of Mycolic Acid Precursors by KasA, a Condensing Enzyme and Drug Target from *Mycobacterium Tuberculosis*. *Journal of Biological Chemistry*. 2013 Nov;288(47):34190–204. doi: 10.1074/jbc.M113.511436.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. 2019 Jan 8;47(Database issue):D607–13.

Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*. 2020 Nov 25;49(D1):D605–12. doi: 10.1093/nar/gkaa1074

Tavares LC. QSAR: a abordagem de Hansch. *Química Nova*. 2004 Aug;27(4):631–9. doi:10.1590/s0100-40422004000400018.

Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455-461. doi:10.1002/jcc.21334

Vilchèze C, Jacobs JR. WR. Resistance to Isoniazid and Ethionamide in *Mycobacterium tuberculosis*: Genes, Mutations, and Causalities. *Microbiology Spectrum*. 2014 Aug 22;2(4).

World Bank. World Bank Country and Lending Groups - World Development Indicators [Internet]. Worldbank.org. 2019. Disponível em: <<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>>. Acesso em 10 ago 2021.

World Health Organization (WHO). Global tuberculosis report 2020: executive summary. Geneva; 2020. Licence: CC BY-NC-SA 3.0 IGO. Disponível em: <<https://www.who.int/publications-detail-redirect/978924001313>>.

## ANEXO I – APROVAÇÃO DO COMITÊ DE PESQUISA



Universidade Federal de Ciências da Saúde de Porto Alegre – UFCSPA

Pró-Reitoria de Pesquisa e Pós-Graduação  
Comissão de Pesquisa

### ATESTADO DE REGISTRO DE PROJETO DE PESQUISA

Atestamos que o projeto de pesquisa intitulado 'Desenvolvimento de Modelos QSAR Para Inibidores da Enzima Enoil ACP Redutase de Mycobacterium tuberculosis' foi registrado na Comissão de Pesquisa da UFCSPA com o número 102/2021, sob responsabilidade de Rafael Andrade Caceres, na forma como foi documentado neste processo 23103.219976/2021-34 até a emissão deste atestado.

O início deste projeto foi marcado em 03 de janeiro de 2022 e o seu término está previsto para 03 de janeiro de 2023.

Salientamos que este registro não autoriza o pesquisador a coletar ou analisar dados oriundos de sujeitos de pesquisa. Salientamos também que este registro não garante a concessão de recursos financeiros por parte da UFCSPA a este projeto de pesquisa.

Comissão de Pesquisa (ComPesq)  
UFCSPA



Documento assinado eletronicamente por Renata Padilha Guedes, Coordenadora de Pesquisa, em 17/12/2021, às 12:52, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufcspa.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufcspa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador 1293391 e o código CRC 92BEF1F3.

## CURRICULO LATES

### Aureo André Karolczak

Curriculum Vitae

---

#### Nome civil

Nome Aureo André Karolczak

#### Dados pessoais

**Filiação** Arnaldo Karolczak e Eloisa Bianchini Karolczak  
**Nascimento** 02/03/1966 - Brasil  
**Carteira de Identidade** 3027263874 ssp - RS - 14/12/1992  
**CPF** 508.707.930-04

**Endereço residencial** Rua Dona Margarida, 64  
Navegantes - Porto Alegre  
93260480, RS - Brasil  
Telefone: 51 32103872  
Celular 51 984463370

**Endereço profissional** Fundação Universidade Federal de Ciências da Saúde de Porto Alegre,  
Programa de Pós-Graduação em Biociências  
Rua Sarmento Leite  
Centro Histórico - Porto Alegre  
90050170, RS - Brasil  
Telefone: 51 33039000

#### Endereço eletrônico

E-mail para contato: andrepoa2002@gmail.com  
E-mail alternativo aureo-karolczak@procergs.rs.gov.br

---

#### Formação acadêmica/titulação

- 2018** Mestrado em BIOCÊNCIAS.  
Fundação Universidade Federal de Ciências da Saúde de Porto Alegre,  
UFCSPA, Porto Alegre, Brasil  
Título: Desenvolvimento de Modelos QSAR para Identificar Inibidores da Enzima enoil-ACP-redutase de *Mycobacterium tuberculosis*  
Orientador: Rafael Andrade Caceres  
Co-orientador: Luis Fernando S. M. Timmers  
*Palavras-chave: Qsar, CoMFA, InhA, Tuberculose*  
*Áreas do conhecimento: Farmacologia Bioquímica e Molecular*  
*Setores de atividade: Fabricação de produtos farmoquímicos e farmacêuticos*
- 2012 - 2018** Graduação em Farmácia.  
Fundação Universidade Federal de Ciências da Saúde de Porto Alegre,  
UFCSPA, Porto Alegre, Brasil  
Título: Desenvolvimento de modelo QSAR para Identificar Inibidores da Enzima Enoil -ACP - Redutase de *Mycobacterium tuberculosis*  
Orientador: Prof. Dr. Rafael Andrade Caceres

**1992 - 1995** Ensino Médio (2o grau).  
Instituto Federal Rio Grande do Sul, IFRS, Brasil, Ano de obtenção: 1995

---

### Formação complementar

**2021 - 2021** Modelagem Molecular em Sistemas Biológicos. (Carga horária: 10h).  
Laboratório Nacional de Computação Científica, LNCC, Petropolis, Brasil  
*Palavras-chave: Sistemas Biológicos*

**2021 - 2021** Curso de curta duração em Dinâmica Molecular Básica. (Carga horária: 10h).  
Laboratório Nacional de Computação Científica, LNCC, Petropolis, Brasil

---

### Atuação profissional

#### 1. PROCERGS - Cia de Processamento de Dados do Estado do Rio Grande do Sul - PROCERGS

##### Vínculo institucional

**1998 - Atual** Vínculo: Servidor público , Enquadramento funcional: Técnico Operacional

---

### Áreas de atuação

1. Teoria da Informação

---

### Idiomas

**Inglês** Compreende Bem, Fala Bem, Escreve Bem, Lê Bem

---

### Totais de produção

#### Eventos

Participações em eventos (encontro)..... 1