UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

MARIANA LOPES GARCIA

Estudos computacionais da enzima N-miristoiltransferase de *Plasmodium falciparum* e seus inibidores como candidatos a agentes antimaláricos

São Carlos
2017

MARIANA LOPES GARCIA

Estudos computacionais da enzima N-miristoiltransferase de *Plasmodium falciparum* e seus inibidores como candidatos a agentes antimaláricos

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada

Opção: Física Biomolecular

Orientador: Prof. Dr. Rafael Victorio Carvalho Guido

Versão Corrigida

(Versão original disponível na Unidade que aloja o Programa)

São Carlos
2017

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica revisada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Garcia, Mariana Lopes

Estudos computacionais da enzima N-miristoiltransferase de *Plasmodium falciparum* e seus inibidores como candidatos a agentes antimaláricos / Mariana Lopes Garcia; orientador Rafael Victorio Carvalho Guido - versão corrigida -- São Carlos, 2017.

105 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2017.

1. Malária. 2. QSAR. 3. N-Miristoiltransferase. I. Guido, Rafael Victorio Carvalho, orient. II. Título.

AGRADECIMENTOS

Primeiramente, agradeço aos meus pais por me incentivarem sempre a seguir minhas escolhas profissionais e por serem exemplos de pessoas honestas e dedicadas.

Agradeço a toda minha família, em especial meu irmão, e à Gabriela Garcia Paschoeto, pelo apoio durante meu trabalho de mestrado.

Ao Prof. Dr. Rafael Victorio Carvalho Guido, pela oportunidade de desenvolver esse trabalho de mestrado e por sempre poder contar com sua ajuda.

Aos meus colegas de laboratório, em especial o Gustavo Machado Lima e a Renata Vieira Bueno, que me ajudaram desenvolver este trabalho.

Ao Instituto de Física de São Carlos, pela oportunidade de realizar o curso de mestrado.

À FAPESP (processo nº 2015/21272-4, Fundação de Amparo à Pesquisa do Estado de São Paulo) e à CAPES pelo auxílio financeiro para desenvolver este trabalho.

*“Em algum lugar, algo incrível
está esperando para ser descoberto”*

Sharon Begley

RESUMO

GARCIA, M. L. **Estudos computacionais da enzima N-miristoiltransferase de *Plasmodium falciparum* e seus inibidores como candidatos a agentes antimaláricos.** 2017. 105 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

A malária é uma doença infecciosa causada pelos parasitas do gênero *Plasmodium* e transmitida pelo mosquito *Anopheles spp.* Devido ao surgimento de casos de resistência aos fármacos disponíveis novos alvos e candidatos a fármacos são necessários. Recentemente, a enzima N-miristoiltransferase (NMT) foi confirmada como essencial para o parasita e validada como alvo terapêutico para o desenvolvimento de candidatos a fármacos antimaláricos. O objetivo desse trabalho foi identificar os determinantes moleculares responsáveis pela atividade inibitória de uma série de derivados benzotiofênicos frente à NMT. Nesse sentido, estudos de relação quantitativa estrutura-atividade (QSAR) 2D e 3D foram desenvolvidos para dois conjuntos de dados de derivados benzotiofênicos como inibidores da enzima do parasita (*Pf*NMT) e a homóloga humana (*Hs*NMT). Além disso, estudos de modelagem por homologia da *Pf*NMT foram conduzidos. Os estudos de QSAR 2D foram desenvolvidos pelo método de Holograma QSAR (HQSAR). O modelo estrutural de *Pf*NMT foi aplicado na construção dos modelos QSAR 3D CoMFA (*Comparative Molecular Field Analysis*) e CoMSIA (*Comparative Molecular Similarity Index Analysis*). Os estudos de QSAR 3D foram conduzidos com diferentes métodos de cálculo de carga parcial atômica (Gasteiger-Hückel, MMFF94 e AM1-BCC, CHELPG e Mulliken) e de alinhamento molecular (Máxima Subestrutura Comum, alinhamento flexível e baseada no alvo molecular). Os melhores modelos construídos pelos métodos de QSAR 2D e 3D foram robustos, internamente consistentes e com elevada capacidade de predição da atividade de novos compostos contra a *Pf*NMT. Os mapas de contribuição e de contorno geraram informações importantes sobre a relação estrutura-atividade dos compostos. Os resultados permitiram a identificação das bases moleculares responsáveis pela atividade dos inibidores benzotiofênicos e são úteis para o planejamento de novos inibidores mais potentes e seletivos para a enzima do parasita.

Palavras-chave: Malária. QSAR. N-Miristoiltransferase.

ABSTRACT

GARCIA, M. L. **Computational studies on *Plasmodium falciparum* N-myristoyltransferase enzyme and its inhibitors as antimalarial drug candidates.** 2017. 105 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

Malaria is an infectious disease caused by protozoan parasites of the genus *Plasmodium* and transmitted by *Anopheles spp.* mosquitos. Due to the emerging resistance to current available drugs, great efforts for new molecular target and drugs are required. Recently, N-myristoyltransferase (NMT) was confirmed as an essential enzyme to malaria parasites and validated as a chemically tractable target for the development of new drug candidates against malaria. This work aimed to shed light on the molecular requirements underlying the inhibitory activity of benzothiophene derivatives against NMT. Therefore, 2D and 3D quantitative structure-activity relationship (QSAR) studies were developed for two datasets of benzothiophene derivatives as *P. falciparum* NMT (*Pf*NMT) and the human homologue (*Hs*NMT) inhibitors. Also, homology modeling studies for *Pf*NMT were developed. The 2D QSAR studies were developed by the Hologram QSAR (HQSAR) method. The *Pf*NMT structural model was applied in the construction of 3D QSAR models CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Index Analysis). Different molecular alignment (maximum common substructure, flexible alignment and structure based) and atomic partial charge calculation (Gasteiger-Hückel, MMFF94, AM1-BCC, CHELPG and Mulliken) methods were used to build the 3D QSAR models. The best models showed internal consistency and high predictive ability of biological activity against *Pf*NMT. The contribution and contour maps gave important information about compounds structure-activity relationship. The results allowed the identification of the molecular requirements underlying the inhibitory activity and should be useful for the design of novel potent and selective *Pf*NMT inhibitors as antimalarial drug candidates.

Keywords: Malaria. QSAR. N-Myristoyltransferase.

LISTA DE FIGURAS

Figura 1 -	Ciclo de vida dos parasitas do gênero <i>Plasmodium</i>	22
Figura 2 -	Mapa de risco por município brasileiro de infecção por malária no ano de 2015	24
Figura 3 -	Países endêmicos para malária em 2000 e em 2016	25
Figura 4 -	Processo de desenvolvimento de novos fármacos	29
Figura 5 -	Construção do holograma molecular no método HQSAR	32
Figura 6 -	Processo de desenvolvimento de um modelo de QSAR 3D CoMFA	35
Figura 7 -	Representação dos campos eletrostáticos e estéreo nos estudos de CoMFA, calculados pelos potenciais de Coulomb e Lennard-Jones, respectivamente	36
Figura 8 -	Representação da função Gaussiana utilizada no cálculo dos descritores moleculares do modelo CoMSIA	37
Figura 9 -	Estrutura molecular da miristoil-coenzima A	38
Figura 10 -	Estruturas 3D da NMT humana (código PDB: 4C2Y). (A) Modelo de fitas. (B) Modelo de superfície com destaque para a cavidade de ligação da molécula de miristoil-CoA	39
Figura 11 -	Estrutura molecular do 1-benzotiofeno	39
Figura 12 -	Representação gráfica da distribuição de pK_i para os conjuntos treinamento, teste e total, mostrando a representatividade dos conjuntos de dados da <i>Pf</i> NMT (A) e <i>Hs</i> NMT (B)	63
Figura 13 -	Representação gráfica dos valores experimentais <i>versus</i> preditos de pK_i para o melhor modelo HQSAR do conjunto de dados da <i>Pf</i> NMT, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos	68
Figura 14 -	Mapas de contribuição dos compostos de menor atividade biológica, à esquerda, e de maior atividade, à direita, obtidos a partir do melhor modelo HQSAR construído com o conjunto de dados da <i>Pf</i> NMT	68
Figura 15 -	Representação gráfica dos valores experimentais <i>versus</i> preditos de pK_i para o melhor modelo HQSAR do conjunto de dados da <i>Hs</i> NMT, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos	72
Figura 16 -	Mapas de contribuição dos compostos de maior atividade biológica (composto 36) e de maior diferença de pK_i (composto 33) entre as enzimas <i>Pf</i> NMT e <i>Hs</i> NMT obtidos a partir dos melhores modelos HQSAR	73
Figura 17 -	Alinhamento das sequências de aminoácidos da NMT de <i>P. falciparum</i> e <i>P. vivax</i> com destaque, em azul, para os resíduos idênticos entre as sequências	74
Figura 18 -	Diagrama de Ramachandran do modelo tridimensional construído da enzima <i>Pf</i> NMT	74
Figura 19 -	Sobreposição da estrutura molde da NMT de <i>P. vivax</i> (PDB: 4CAF), em azul, e o modelo construído da <i>Pf</i> NMT, em verde. (A) Estruturas totais e (B) Sítio de ligação dos compostos benzotiofênicos após modificação da conformação dos resíduos marcados, com o composto 34 cristalográfico representado em magenta	75

Figura 20 -	Representação da estrutura molecular do composto 36 na conformação de menor energia para os diferentes tipos de carga atômica parcial calculados. (A) Carga Gasteiger-Hückel; (B) Carga MMFF94; (C) Carga AM1-BCC; (D) Carga CHELPG; (E) Carga Mulliken. O espectro de cores (azul, branco e vermelho) representam a distribuição de cargas parciais atômicas do composto	76
Figura 21 -	Representação do alinhamento estrutural baseado na MCS dos compostos na conformação de menor energia para os diferentes tipos de carga atômica parcial calculados. (A) Carga Gasteiger-Hückel; (B) Carga MMFF94; (C) Carga AM1-BCC; (D) Carga CHELPG; (E) Carga Mulliken	77
Figura 22 -	Representação do alinhamento flexível de ligantes dos compostos para os diferentes tipos de carga atômica parcial calculados. (A) Carga Gasteiger-Hückel; (B) Carga MMFF94; (C) Carga AM1-BCC; (D) Carga CHELPG; (E) Carga Mulliken	78
Figura 23 -	Representação dos alinhamentos estruturais baseado na estrutura do alvo molecular (<i>Pf</i> NMT) desenvolvidos pela docagem molecular dos compostos com os programas (A) Surflex-Dock e (B) Glide	79
Figura 24 -	Representação gráfica dos valores experimentais <i>versus</i> preditos de pK_i para os melhores modelos (A) CoMFA e (B) CoMSIA do conjunto de dados da <i>Pf</i> NMT, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos	86
Figura 25 -	Mapas de contorno gerados para o melhor modelo CoMFA a partir do conjunto de dados da <i>Pf</i> NMT. (A) Mapa de contorno eletrostático, representado em azul e vermelho (níveis de contorno 0,10 e -0,05 kcal/mol, respectivamente). (B) Mapa de contorno estérico, representado em verde e amarelo (níveis de contorno 0,005 e -0,001 kcal/mol, respectivamente)	87
Figura 26 -	Mapas de contorno gerados para o melhor modelo CoMSIA a partir do conjunto de dados da <i>Pf</i> NMT. (A) Mapa de contorno hidrofóbico representado em amarelo e branco (níveis de contorno 0,005 e -0,010 kcal/mol, respectivamente). (B) Mapa de contorno eletrostáticos representado em azul e vermelho (níveis de contorno 0,003 e -0,003 kcal/mol, respectivamente). (C) Mapa de contorno de doadores, representado em ciano e roxo (níveis de contorno 0,01 e -0,05 kcal/mol, respectivamente) e aceptores de ligação de hidrogênio, representado em magenta e amarelo (níveis de contorno 0,05 e -0,03 kcal/mol, respectivamente)	87
Figura 27 -	Representação gráfica dos valores experimentais <i>versus</i> preditos de pK_i para os melhores modelos (A) CoMFA e (B) CoMSIA do conjunto de dados da <i>Hs</i> NMT, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos	92
Figura 28 -	Mapas de contorno gerados para os melhores modelos CoMFA a partir do conjunto de dados da <i>Pf</i> NMT (à esquerda) e <i>Hs</i> NMT (à direita), sendo o composto mais potente (36) a molécula de referência. (A) Mapa de contorno estérico (B) Mapa de contorno eletrostático	93
Figura 29 -	Mapas de contorno gerados para os melhores modelos CoMSIA a partir do conjunto de dados da <i>Pf</i> NMT (à esquerda) e <i>Hs</i> NMT (à direita), sendo o composto mais potente (36) a molécula de referência. (A) Mapa de contorno hidrofóbico. (B) Mapa de contorno eletrostáticos. (C) Mapa de contorno de doadores e aceptores de ligação de hidrogênio	94
Figura 30 -	Representação das modificações estruturais sugeridas para o composto 36 pelos modelos HQSAR, CoMFA e CoMSIA	96

LISTA DE TABELAS

Tabela 1 -	Compostos utilizados como fármacos antimaláricos	26
Tabela 2 -	Parâmetros de distinção de fragmento no método HQSAR	33
Tabela 3 -	Estruturas químicas dos derivados 1-benzotiofenos com atividade <i>in vitro</i> frente à NMT de <i>P. falciparum</i> e <i>H. sapiens</i>	48
Tabela 4 -	Resultados estatísticos dos modelos HQSAR iniciais para os inibidores benzotiofênicos contra a enzima <i>Pf</i> NMT, com tamanho de fragmento de 4 – 7 átomos	64
Tabela 5 -	Resultados estatísticos dos modelos HQSAR para os inibidores benzotiofênicos contra a enzima <i>Pf</i> NMT com a retirada do composto <i>outlier</i> 7 e tamanho de fragmento de 4 – 7 átomos	65
Tabela 6 -	Resultados estatísticos dos modelos HQSAR dos inibidores benzotiofênicos contra a enzima <i>Pf</i> NMT, variando o tamanho dos fragmentos e mantendo a distinção de fragmento A/B/C/H/DA	66
Tabela 7 -	Valores experimentais e preditos de pK _i dos compostos do conjunto teste obtidos pelo melhor modelo HQSAR da <i>Pf</i> NMT	66
Tabela 8 -	Resultados estatísticos dos modelos HQSAR iniciais para os inibidores benzotiofênicos contra a enzima <i>Hs</i> NMT, com tamanho de fragmento de 4 – 7 átomos	69
Tabela 9 -	Resultados estatísticos dos modelos HQSAR para os inibidores benzotiofênicos contra a enzima <i>Hs</i> NMT com a retirada dos compostos <i>outliers</i> 6 e 10 e tamanho de fragmento de 4 – 7	70
Tabela 10 -	Resultados estatísticos dos modelos HQSAR dos inibidores benzotiofênicos contra a enzima <i>Hs</i> NMT, variando o tamanho dos fragmentos e mantendo a distinção de fragmento A/B/C/DA	71
Tabela 11 -	Valores experimentais e preditos de pK _i dos compostos do conjunto teste obtidos pelo melhor modelo HQSAR da <i>Hs</i> NMT	72
Tabela 12 -	Resultados estatísticos dos modelos QSAR 3D CoMFA iniciais para os inibidores benzotiofênicos contra a enzima <i>Pf</i> NMT	80
Tabela 13 -	Resultados estatísticos dos modelos QSAR 3D CoMSIA iniciais para os inibidores benzotiofênicos contra a enzima <i>Pf</i> NMT	81
Tabela 14 -	Análise dos modelos CoMFA com base nos parâmetros estatísticos para diferentes definições de espaçamento do <i>grid</i> e região de focagem para o conjunto de dados da <i>Pf</i> NMT	83
Tabela 15 -	Valores experimentais e preditos de pK _i dos compostos do conjunto teste obtidos pelo melhor modelo CoMFA da <i>Pf</i> NMT	84
Tabela 16 -	Valores experimentais e preditos de pK _i dos compostos do conjunto teste obtidos pelo melhor modelo CoMSIA da <i>Pf</i> NMT	85
Tabela 17 -	Resultados estatísticos dos modelos QSAR 3D, CoMFA e CoMSIA, para os inibidores benzotiofênicos contra a enzima <i>Hs</i> NMT	89

Tabela 18 -	Análise dos modelos CoMFA com base nos parâmetros estatísticos para diferentes definições de espaçamento do <i>grid</i> e região de focagem para o conjunto de dados da <i>HsNMT</i>	90
Tabela 19 -	Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo CoMFA da <i>HsNMT</i>	91
Tabela 20 -	Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo CoMSIA da <i>HsNMT</i>	91

LISTA DE ABREVIATURAS E SIGLAS

ACT	Terapia Combinada com Artemisinina
CIBFar	Centro de Pesquisa e Inovação em Biodiversidade e Fármacos
CEPID	Centros de Pesquisa, Inovação e Difusão
CoMFA	Análise Comparativa dos Campos Moleculares (<i>Comparative Molecular Field Analysis</i>)
CoMSIA	Análise Comparativa dos Índices de Similaridades Moleculares (<i>Comparative Molecular Similarity Indices Analysis</i>)
DNA	Ácido Desoxirribonucleico (Deoxyribonucleic acid)
hERG	Canal iônico humano ERG (<i>human Ether-a-go-go-Related Gene</i>)
HIV	Vírus da Imunodeficiência Humana (<i>Human Immunodeficiency Virus</i>)
HQSAR	Holograma Relação Quantitativa de Estrutura-Atividade (<i>Hologram Quantitative Structure-Activity Relationship</i>)
HsNMT	N-Miristoiltransferase de <i>Homo sapiens</i>
K _i	Constante de inibição
LBDD	Planejamento de Fármacos Baseado no Ligante (<i>Ligand-Based Drug Discovery</i>)
LOO	Método de validação cruzada <i>Leave-one-out</i>
MCS	Máxima Subestrutura Comum (<i>Maximum Common Substructure</i>)
NMT	N-Miristoiltransferase
OMS	Organização Mundial da Saúde
PfNMT	N-Miristoiltransferase de <i>Plasmodium falciparum</i>
pK _i	Logaritmo da constante de inibição
PLS	Análise de Regressão dos Quadrados Mínimos Parciais (<i>Partial Least Square</i>)
PNCM	Programa Nacional de Controle de Malária
PvNMT	N-Miristoiltransferase de <i>Plasmodium vivax</i>
P&D	Pesquisa e Desenvolvimento
QSAR	Relação Quantitativa de Estrutura-Atividade (<i>Quantitative Structure-Activity Relationship</i>)
RMS	Raiz do Valor Quadrático Médio (<i>Root Mean Square</i>)
RMSD	Desvio Quadrático Médio (<i>Root Mean Square Deviation</i>)
SAR	Relação Estrutura-Atividade (<i>Structure-Activity Relationship</i>)

SBDD	Planejamento de Fármacos Baseado na Estrutura do Alvo (<i>Structure-Based Drug Discovery</i>)
SUS	Sistema Único de Saúde

SUMÁRIO

1 INTRODUÇÃO	19
1.1 Malária	21
1.1.1 Agente etiológico, Ciclo de vida e Patologia	21
1.1.2 Epidemiologia	23
1.1.3 Fármacos antimaláricos e o desenvolvimento de resistência	25
1.2 Desenvolvimento Racional de Novos Fármacos	28
1.3 Relações Quantitativas de Estrutura e Atividade	30
1.3.1 Holograma QSAR	32
1.3.2 Análise Comparativa dos Campos Moleculares (CoMFA)	34
1.3.3 Análise Comparativa dos Índices de Similaridade Molecular (CoMSIA)	36
1.4 Planejamento de Novos Antimaláricos	37
1.4.1 N-Miristoiltransferase	38
1.4.2 Compostos derivados de benzotiofeno	39
2 OBJETIVOS	41
3 ESTRATÉGIAS E MÉTODOS	45
3.1 Estratégias computacionais	47
3.2 Criação de conjunto de dados para modelagem QSAR	47
3.3 Modelagem QSAR: QSAR 2D e QSAR 3D	55
3.3.1 QSAR 2D – Holograma QSAR	55
3.3.2 QSAR 3D – CoMFA e CoMSIA	56
3.3.2.1 Cargas parciais atômicas	56
3.3.2.2 Alinhamento molecular	56

3.3.2.3 Construção dos modelos	58
3.4 Modelagem por homologia	58
4 RESULTADOS E DISCUSSÕES	61
4.1 Preparo do conjunto de dados	63
4.2 Construção dos modelos HQSAR	63
4.2.1 Modelos HQSAR para o conjunto de dados da <i>Pf</i> NMT	63
4.2.2 Modelos HQSAR para o conjunto de dados da <i>Hs</i> NMT	68
4.3 Construção do modelo estrutural da enzima <i>Pf</i>NMT	73
4.4 Construção dos modelos QSAR 3D	75
4.4.1 Cálculo de cargas atômicas parciais	76
4.4.2 Alinhamento molecular	77
4.4.3 Modelos CoMFA e CoMSIA para o conjunto de dados da <i>Pf</i> NMT	79
4.4.4 Modelos CoMFA e CoMSIA para o conjunto de dados da <i>Hs</i> NMT	88
4.5 Planejamento de novos candidatos a fármacos antimaláricos	95
5 CONCLUSÕES E PERSPECTIVAS	97
REFERÊNCIAS	101

INTRODUÇÃO

1 INTRODUÇÃO

1.1 Malária

1.1.1 Agente etiológico, Ciclo de vida e Patologia

Malária é a doença infecciosa parasitária causada pelos protozoários do gênero *Plasmodium* e transmitida pela picada da fêmea do mosquito *Anopheles spp.* Cinco espécies são agentes etiológicos da doença em humanos: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* e *P. knowlesi*. A forma grave da doença pode ser causada por qualquer uma das espécies, no entanto, a espécie *P. falciparum* é responsável por quase todos esses casos e, consequentemente, pela maioria das mortes por malária.¹⁻²

Os parasitas da malária apresentam um complexo ciclo de vida dividido em dois estágios: assexuado, no hospedeiro humano, e sexuado, desenvolvido no mosquito (Figura 1). A infecção em humanos ocorre após a transmissão do parasita na forma de esporozoítos da saliva do mosquito para a corrente sanguínea do hospedeiro. Os esporozoítos migram para o fígado e invadem os hepatócitos, onde podem se incubar ou se multiplicar para formar esquizontes, que depois se rompem liberando milhares de parasitas na forma assexuada, chamada merozoíto. Os merozoítos invadem os eritrócitos e se transformam em trofozoítos, que crescem e se reproduzem dando origem a novos merozoítos (de 8 a 32), o que causa a ruptura da célula. Algumas poucas formas assexuadas haploides se diferenciam em gametócitos masculino e feminino dentro dos eritrócitos. Os gametócitos são ingeridos pelo mosquito em uma nova picada e, então, diferenciam-se em gametas que se fundem e formam um zigoto no intestino médio do mosquito. O zigoto sofre meiose dando origem a oocinetos capazes de se mover espontaneamente e atravessar a parede do intestino. Os oocinetos eventualmente formam oocistos que passam por processo de replicação de DNA e originam milhares de esporozoítos. Esses esporozoítos migram para a glândula salivar do mosquito e podem ser transmitidos novamente para o próximo hospedeiro humano a ser picado.³⁻⁴

As espécies *P. vivax* e *P. ovale* apresentam um estágio hepático latente (hipnozoítos), na qual o parasita permanece dormente nas células por meses ou até anos, podendo causar os sintomas no hospedeiro diversas vezes. Infecções de malária causadas por essas espécies necessitam de um tratamento especial que vise eliminar o estágio de hipnozoíto do parasita.¹

Figura 1 - Ciclo de vida dos parasitas do gênero *Plasmodium*

Fonte: Adaptada de FLANNERY³

Os ciclos de invasão, replicação e destruição dos eritrócitos realizados pelos merozoítos resultam no crescimento exponencial da população de parasitas e no desenvolvimento da doença. Cerca de 66% da hemoglobina contida nos eritrócitos é consumida em vacúolos digestivos do parasita *P. falciparum* como a maior fonte de nutrientes.⁴ A ruptura dos eritrócitos aderidas às células endoteliais de tecidos e órgãos liberam citocinas pró-inflamatórias que ativam o sistema imune inato. Como consequência, o indivíduo infectado sofre uma inflamação sistêmica que precede o aparecimento dos primeiros sintomas entre 7 a 15 dias após a infecção. Entre esses sintomas estão febre, dor de cabeça, dores musculares, calafrios, vômitos e letargia. Além disso, a doença é caracterizada pelo que é chamado de paroxismo malárico, no qual o indivíduo infectado passa por ataques súbitos dos sintomas que se repetem diversas vezes, o que está relacionado aos ciclos de invasão dos eritrócitos.⁵⁻⁶

Quadros sintomáticos simples de malária causada pelo *P. falciparum* podem evoluir rapidamente para formas mais graves da doença se não tratada. As principais síndromes associadas à malária grave são a anemia grave, acidose metabólica e malária cerebral.⁴ A malária cerebral, caracterizada pelo acometimento do sistema nervoso central, causa convulsão seguida pelo coma e é a principal causa de óbito por malária. Os mecanismos que causam lesões cerebrais pelos parasitas ainda não são totalmente conhecidos.⁷

1.1.2 Epidemiologia

A malária é a infecção parasitária que mais causa mortes no mundo. As manifestações de malária nas regiões globais são determinadas por fatores intrínsecos - imunidade humana, espécies de parasitas, longevidade do mosquito etc - e fatores extrínsecos - condições ambientais, condições econômicas e políticas, efetividade de programas de controle e prevenção da doença.⁸ Países de regiões tropicais e subtropicais são mais afetados por casos de malária, dado o clima quente e úmido propício para a reprodução do mosquito vetor *Anopheles*. Além disso, a quantidade de casos de malária tem relação com o índice de desenvolvimento humano da região analisada, sendo que as populações de países subdesenvolvidos, que têm pouco ou nenhum acesso a informações de prevenção, tratamentos médicos e medicamentos, sofrem mais com infecções parasitárias.⁹

Crianças e mulheres grávidas apresentam maior mortalidade pelas infecções maláricas. Mulheres grávidas são mais susceptíveis à infecção devido à baixa imunidade. Nos casos de malária na gravidez, o desenvolvimento neurossensorial, cognitivo e comportamental do bebê pode ser prejudicado pelo nascimento prematuro. Para as crianças, a alta taxa de mortalidade se deve a ser a primeira vez em que o indivíduo é infectado por malária, ou seja, ainda não apresenta resposta imune contra o parasita.¹⁰

No ano de 2015, a Organização Mundial da Saúde (OMS) estimou 212 milhões de casos de malária no mundo, sendo 90% ocorridos em países africanos, 7% em países do sul e leste asiático e 2% na região leste mediterrânea. Desses casos, estima-se que 429 mil pessoas morreram da doença, sendo a vasta maioria das infecções (99%) causada pelo parasita *P. falciparum*. Dentre os casos de morte por malária no mundo, 70% corresponde a crianças menores de 5 anos de idade. O *P. falciparum* é o parasita predominante nos países do continente africano, enquanto o *P. vivax* é o mais comum na maioria dos países fora da África Subsaariana, o que explica a grande quantidade de óbitos em países africanos.¹¹

No Brasil, a região Amazônica é considerada endêmica para a malária, sendo que os casos ocorrem nos estados do Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. Em 2016, foram reportados 128,7 mil casos de malária na região amazônica do Brasil, sendo a maioria nos estados Amazonas (42%), Acre (26%) e Amapá (10%). Dentre esses casos, aproximadamente 10% foram causados pela espécie *P. falciparum*. A Figura 2 apresenta o mapa de risco de infecção por malária no ano de 2015, que mostra a grande concentração de risco na região norte do país (Fonte: Sivep-Malária/SVS - Ministério da Saúde).

O total de financiamentos mundiais recebidos para o controle e eliminação da malária no ano de 2015 foi estimado em 2,9 bilhões de dólares, apresentando um crescimento de 60 milhões se comparado ao ano de 2010. Entre as fontes desses recursos financeiros estão principalmente governos internacionais, fundações privadas de caridade como a Fundação Bill e Melinda Gates e instituições da União Europeia, sendo esses recursos direcionados para o controle e eliminação da malária pelo Fundo Global para o Combate de AIDS, Tuberculose e Malária.¹¹

Entre os anos de 2000 e 2015, a quantidade de casos de malária caiu 41%, assim como a taxa de mortalidade, que teve uma diminuição de 62% no mundo. A Figura 3 apresenta os 17 países que deixaram de ser endêmicos para a malária entre os anos 2000 e 2016.¹¹

Figura 2 - Mapa de risco por município brasileiro de infecção por malária no ano de 2015
Fonte: Adaptada de BRASIL ...⁹⁰

A redução nos casos de malária é atribuída ao conjunto de medidas estratégicas para o controle da doença. Entre essas medidas estão o uso de mosquiteiros tratados com inseticida, *sprays* anti-inseto, terapia preventiva intermitente para grávidas, melhores diagnósticos e efetividade das terapias combinadas com artemisinina (ACTs). Como explicação para as regiões onde os casos de malária se mantiveram ou aumentaram, pode-se considerar que as intervenções contra a doença não foram suficientes ou não tiveram efetividade, devido principalmente à resistência aos fármacos utilizados.¹²

Figura 3 - Países endêmicos para malária em 2000 e em 2016

Fonte: Adaptada de WORLD ...¹¹

1.1.3 Fármacos antimaláricos e o desenvolvimento de resistência

A busca por novos compostos com atividade antimalárica sempre será essencial no combate contra a malária. O surgimento e disseminação de resistência aos fármacos existentes são uma grande preocupação para as doenças infecciosas. Diversos medicamentos já foram utilizados no tratamento de malária na história, no entanto, muitos deles não são mais adequados devido ao desenvolvimento de resistência pelos parasitas.¹³

Atualmente, os tratamentos para a malária recomendados pela OMS são principalmente os ACTs, nos quais são combinados um composto derivado de artemisinina (arteméter, artesunato, diidroartemisinina) e um outro composto ao qual o parasita não seja resistente, como a lumefantrina, amodiaquina, mefloquina, piperquina e sulfadoxina-pirimetamina. A cloroquina é utilizada em regiões onde os parasitas ainda são susceptíveis a esse composto. Mulheres grávidas, que não podem receber o tratamento com artemisinina, são tratadas com quinina mais clindamicina. A primaquina é utilizada para a prevenção da reincidência da doença, por ser ativo contra a forma de hipnozoítos dos parasitas *P. vivax* e *P. ovale*.¹⁴ A principal vantagem do uso de artemisininas é a velocidade de eliminação de novos parasitas na fase intraeritrocítica de anel, o que previne a posterior maturação.¹⁵ A Tabela 1 apresenta as classes e estruturas químicas desses compostos.

No Brasil, o Ministério da Saúde orienta o tratamento e disponibiliza os medicamentos contra a malária pelo Sistema Único de Saúde (SUS), o que faz parte do Programa Nacional de

Controle de Malária (PNCM). A decisão de qual tratamento será utilizado se baseia em informações de diagnóstico que incluem: a espécie de plasmódio infectante, idade do paciente, histórico de infecções por malária, condições de risco (gravidez, portador de HIV etc.) e gravidade da doença (Guia Prático de Tratamento da Malária no Brasil <<http://portalarquivos.saude.gov.br/images/pdf/2014/junho/11/Guia-pr--tico-de-tratamento-da-mal--ria-no-Brasil.pdf>>)

Tabela 1 - Compostos utilizados como fármacos antimaláricos

Composto	Classe	Estrutura química
Quinina	Quinolinometanóis
Mefloquina	Quinolinometanóis
Cloroquina	4-Aminoquinolinas
Amodiaquina	4-Aminoquinolinas
Piperaquina	4-Aminoquinolinas
Primaquina	8-Aminoquinolinas
Pirimetamina	Antifolatos

continua

continuação

Tabela 1 - Compostos utilizados como fármacos antimaláricos (continuação)

Composto	Classe	Estrutura química
Sulfadoxina	Antifolatos
Lumefantrina	Aril-álcoois
Diidroartemisinina	Derivados de artemisinina
Artesunato	Derivados de artemisinina
Arteméter	Derivados de artemisinina

Fonte: Elaborada pela autora

A maior parte dos fármacos antimaláricos têm como alvo a fase assexuada eritrocítica dos plasmódios.¹⁶ Durante esse estágio, os parasitas degradam hemoglobina em vacúolos digestivos para conseguir nutrientes, o que gera como produto grupos heme livres que reagem com oxigênio molecular formando a hematina. A hematina é tóxica para os parasitas, causa danos na membrana celular e lise. Para se livrar da toxicidade dos produtos liberados, os parasitas convertem dímeros de hematina em cristais inertes, a hemozoína.⁴ Comumente, os fármacos antimaláricos agem desestabilizando a polimerização de hematina e intoxicando os parasitas com os próprios produtos da destruição da hemoglobina.

Parasitas resistentes podem surgir devido ao uso excessivo de fármacos antimaláricos para profilaxia ou tratamentos incompletos de pacientes com malária, por descumprimento do

tratamento adequado, ou por características das populações do parasita, como flexibilidade genética e metabólica, alta taxa de reprodução e disseminação dos genes de resistência pela recombinação sexual no mosquito.¹⁷ Os plasmódios já apresentam alguma forma de resistência para todos os fármacos citados anteriormente, sendo que os mais recentes casos de resistência foram desenvolvidos contra os derivados de artemisinina e as terapias combinadas com esse fármaco.^{15,18}

1.2 Desenvolvimento Racional de Novos Fármacos

O processo de descoberta de um fármaco é altamente complexo e dispendioso, envolvendo requisitos de inovação, conhecimento e investimentos em pesquisa e desenvolvimento (P&D).¹⁹⁻²⁰ Com a revolução biotecnológica gerada pelos recentes estudos geno-, proteo- e metabolômicos, a descoberta de novos fármacos evolui do modelo tradicional de “tentativa e erro” para um processo mais racional. O aumento de expectativa de vida de diversas populações é consequência desse grande desenvolvimento científico que culminou na descoberta de novas gerações de fármacos para o tratamento de diversas doenças humanas.²⁰⁻²² A química medicinal, área científica de caráter multidisciplinar, combina conceitos de química orgânica, bioquímica e biologia molecular e estrutural, tornando-se a base científica necessária para a implementação de estratégias inovadoras na descoberta de fármacos.²³⁻²⁴

Os estudos para descoberta e desenvolvimento de um novo fármaco podem ser divididos em duas fases: fase pré-clínica e fase clínica (Figura 4). Durante a fase pré-clínica, inicialmente é identificada uma necessidade médica a qual seja possível e adequado o desenvolvimento de um novo fármaco. Então, é analisado como essa doença alvo pode ser tratada e, se já existem terapias para essa doença, como elas podem ser melhoradas. Novos compostos são testados na busca de um composto *hit* – molécula biologicamente ativa que se destaque por sua maior potência dentre as outras em um dado ensaio.^{20,25} A partir do composto *hit*, inicia-se a fase *hit-to-lead*, que tem como objetivo produzir compostos mais potentes e seletivos com propriedades farmacocinéticas adequadas para terem a eficácia *in vivo* avaliada.²⁶ Durante essa fase são desenvolvidos os estudos de relação estrutura-atividade (SAR) e também podem ser feitos estudos baseados na estrutura do alvo molecular, caso seja conhecido. Na fase seguinte, denominada lead optimization (otimização do composto líder), o composto sofre modificação na estrutura química buscando a manutenção da potência, otimização das propriedades farmacocinética e menor interação com o canal iônico hERG.²⁷ Além disso, testes

toxicológicos em animais são exigidos antes de os compostos serem aprovados para a posterior fase clínica.²⁵

Figura 4 - Processo de desenvolvimento de novos fármacos
Fonte: Adaptada de LOMBARDINO; LOWE.²⁵

A fase clínica do processo de desenvolvimento de um novo fármaco apresenta quatro fases. Na Fase 1, o composto é testado em humanos saudáveis para examinar a tolerância ao medicamento. Na Fase 2, são investigados a eficácia e as doses necessárias, administrando o composto em pacientes com a doença alvo. A Fase 3 faz testes em milhares de pacientes para criar uma ampla base de dados de eficácia e segurança. Apenas 4-7% dos candidatos a fármacos conseguem passar por esses testes e são aprovados como medicamentos pelo órgão regulador.²⁵ Por fim, a Fase 4, após a comercialização do fármaco, inclui estudos de farmacovigilância que visam a identificação de efeitos adversos não observados nas fases anteriores de desenvolvimento.

O desenvolvimento racional de novos medicamentos busca utilizar todo o conhecimento disponível sobre o sistema em questão para a criação de uma estratégia de otimização dos compostos biologicamente ativos.²⁸ A aplicação de métodos computacionais tem se mostrado muito útil no processo de desenvolvimento de fármacos desde a fase de identificação de compostos *hit* até a fase de otimização do composto líder, economizando tempo e custos com sínteses químicas e testes biológicos.²⁹ A descoberta de uma grande quantidade de moléculas bioativas foi possível devido ao uso de métodos computacionais de planejamento baseado na estrutura do ligante LBDD (do inglês, *Ligand-Based Drug Design*) ou planejamento baseado na estrutura do receptor SBDD (do inglês, *Structure-Based Drug Design*).³⁰⁻³¹

Os métodos de SBDD utilizam informação sobre a topologia do alvo molecular obtida por cristalografia de raios X, Ressonância Magnética Nuclear ou pela construção de um modelo teórico robusto (modelagem por homologia, *ab initio*, etc). Nesse contexto, ferramentas computacionais são empregadas para avaliar a complementaridade entre diversas moléculas e a cavidade de ligação do alvo biológico. Modelos farmacofóricos são construídos com base nas possíveis interações moleculares no interior da cavidade. No caso de doenças parasitárias, os modelos farmacofóricos são especialmente atrativos por serem comumente utilizados para explorar as diferenças estruturais entre o alvo molecular do parasita e do homólogo humano, possibilitando um estudo de seletividade para o desenvolvimento de novos compostos bioativos.³¹

O processo de LBDD inicia-se pelo estudo das características multifuncionais dos ligantes conhecidos ou do ligante endógeno envolvido na fisiopatologia/farmacologia do processo sob investigação. As principais estratégias de LBDD são fundamentadas na análise comparativa de moléculas bioativas, entre elas destacam-se os estudos de similaridade molecular, construções farmacofóricas e de relação quantitativa estrutura-atividade (QSAR, do inglês, *Quantitative Structure-Activity Relationships*), que buscam identificar e quantificar as relações predominantes entre a estrutura química e a atividade biológica de um conjunto de moléculas quimicamente relacionadas por meio do uso de descritores químicos.²⁰

1.3 Relações Quantitativas de Estrutura-Atividade

Desde a criação do método QSAR por Hansch & Fujita³² em 1963, os estudos de QSAR têm sido aplicados na construção de modelos estatísticos robustos importantes no processo de desenvolvimento de novos fármacos.³³ Esses modelos relacionam os descritores moleculares dos ligantes e a respectiva atividade biológica determinada experimentalmente, para fazer previsões de atividade de novas moléculas, mostrar quais características estruturais são mais influentes na atividade e otimizar o desenvolvimento de compostos potentes.³⁴

Para construir um modelo QSAR, é necessário extrair descritores da estrutura molecular que sejam informativos na análise de atividade biológica. Os valores desses descritores são utilizados como variáveis independentes na construção do modelo, relacionados com a atividade biológica (variável dependente) por meio de técnicas estatísticas de correlação de dados.³⁵

Os métodos de QSAR são mais comumente diferenciados com base nos descritores moleculares que caracterizam a estrutura dos compostos do conjunto de dados. Os tipos de

descritores moleculares podem ser categorizados em: QSAR 1D, relação da atividade biológica com uma propriedade geral da molécula como pK_a , $\log P$ etc.; QSAR 2D, relação da atividade com padrões estruturais como índices de conectividade, holograma molecular, farmacóforos 2D etc.; QSAR 3D, relação da atividade com campos de interação não-covalente que cercam a molécula; QSAR 4D, adiciona um conjunto de conformações do ligante ao QSAR 3D; QSAR 5D, acrescenta ao método QSAR 4D diferentes representações do modelo com ajuste induzido; QSAR 6D, incorpora diferentes modelos de solvatação ao método QSAR 5D.³⁶

Os métodos de QSAR também podem ser classificados de acordo com o tipo de técnica de correlação utilizada para estabelecer a relação entre estrutura e atividade. As técnicas podem ser lineares, como a regressão linear (LR), regressão linear múltipla (MLR), mínimos quadrados parciais (PLS) e análise/regressão de componentes principais (PCA/PCR), ou podem ser não-lineares, como redes neurais artificiais (ANN), k -vizinhos mais próximos (kNN) e rede neural com aproximação bayesiana.³⁶

A construção de modelos QSAR com consistência interna e com alta capacidade preditiva requer a análise de alguns fatores. A qualidade do conjunto de dados é essencial nesse contexto, os valores de atividade biológica devem ter sido determinados por experimentos validados e padronizados para todos os compostos, que devem agir pelo mesmo mecanismo.³⁷⁻³⁸ O conjunto de dados deve apresentar uma distribuição homogênea de valores de atividade biológica e que abranjam pelo menos 3 ordens logarítmicas. Os compostos devem ser divididos em conjunto treinamento (70-80% das moléculas), para construção do modelo, e conjunto teste (20-30% das moléculas), para validação externa.³³ Além disso, é necessário que o conjunto treinamento contenha pelo menos 20 moléculas e ambos conjuntos devem apresentar variabilidade estrutural e de atividade biológica, sendo representativos do conjunto total.³⁷

Os descritores de QSAR 2D têm em comum a característica de independência da orientação tridimensional dos compostos do conjunto de dados.³⁹ Esses descritores podem ser constitucionais, eletrostáticos/quânticos, topológicos, geométricos, baseados em fragmentos moleculares ou *molecular fingerprint* (impressão digital molecular). O Holograma QSAR (HQSAR) é um método de QSAR 2D que utiliza descritores do tipo impressão digital molecular, obtidos mais rapidamente em relação a outros métodos e com resultados comparáveis aos dos métodos de QSAR 3D.^{35,40}

Os métodos de QSAR 3D se baseiam na hipótese de que a atividade biológica depende do arranjo tridimensional (conformação) da molécula.⁴¹ Os métodos de QSAR 3D podem ser do tipo baseada no ligante, que usam informações estruturais apenas dos compostos biologicamente ativos na construção do modelo, ou baseado no receptor, que usa informações

do ligante e também do receptor molecular, quando esta estrutura é disponível. Além disso, podem ser classificadas como dependentes ou independentes de alinhamento molecular.³⁶ Os métodos de QSAR 3D mais populares são a Análise Comparativa de Campos Moleculares (CoMFA, do inglês *Comparative Molecular Field Analysis*) e Análise Comparativa de Índices de Similaridade Molecular (CoMSIA, do inglês *Comparative Molecular Similarity Index Analysis*), classificados como baseados no ligante e dependentes de alinhamento molecular.³⁰

1.3.1 Holograma QSAR

Holograma QSAR (HQSAR)⁴² é um método de QSAR 2D fundamentado na construção dos hologramas moleculares, que são obtidos a partir da fragmentação molecular bidimensional. Esse método não requer nenhuma informação tridimensional explícita (como a determinação de conformações e do alinhamento molecular) ou cálculo de descritores diversos, sendo mais simples que os métodos de QSAR 3D e clássicos.⁴³

O método se divide em duas etapas: o cálculo dos descritores moleculares e posterior análise estatística. A etapa de cálculo dos descritores moleculares corresponde à construção dos hologramas moleculares, ilustrado na Figura 5. Para a construção dos hologramas, a estrutura de cada molécula é fragmentada gerando diversos tipos de fragmentos (lineares, ramificados, cíclicos e sobrepostos). A cada fragmento específico é atribuído um número positivo inteiro pelo algoritmo cíclico de verificação de redundância (CRC, do inglês, *cyclic redundancy check*). Em seguida, cada um desses números gerados é alocado em uma posição do holograma molecular pelo algoritmo de *hashing*.

Figura 5 - Construção do holograma molecular no método HQSAR
Fonte: Adaptada de LOWIS⁴²

As ocupâncias de cada posição são incrementadas quando o mesmo tipo de fragmento é contado. No entanto, como o algoritmo de *hashing* causa a diminuição do comprimento dos hologramas, pode ocorrer um fenômeno conhecido como colisão de fragmentos, devido ao número de posições no holograma ser menor do que o número de fragmentos específicos gerados. A colisão de fragmentos ocorre quando fragmentos distintos são representados na mesma posição do holograma. Dessa forma, é essencial analisar diferentes tamanhos de holograma durante a geração de modelos HQSAR para garantir que os resultados não são consequência de colisão de fragmentos, dado que a variação desse parâmetro causa a mudança de ocupância das posições.⁴⁰

A construção dos hologramas moleculares depende de três parâmetros que são definidos pelo usuário: tamanho dos fragmentos (de N a M átomos por fragmento), comprimento do holograma (HL) e distinção dos fragmentos, que pode ser baseada em: átomos (A), ligações (B), conectividade (C), hidrogênios (H), quiralidade (Ch) e doadores e aceptores de ligação de hidrogênio (DA).⁴⁰ A Tabela 2 mostra como a distinção de fragmento é implementada na construção dos hologramas. Esses parâmetros ditam como a estrutura molecular será descrita na forma de holograma, sendo assim, variar as combinações de parâmetros pode gerar modelos de maior ou menor capacidade preditiva.

Tabela 2 - Parâmetros de distinção de fragmento no método HQSAR

Tipo de distinção	Definição
Átomos (A)	Tipos de átomos
Ligações (B)	Tipos de ligação química (simples, dupla, tripla, aromática)
Conectividade (C)	Hibridação dos átomos
Hidrogênios (H)	Presença e número de átomo de hidrogênio
Quiralidade (Ch)	Presença de centros quirais
Doadores e Aceptores de Ligação de Hidrogênio (DA)	Presença de átomos aceptores ou doadores de ligação de hidrogênio

Fonte: Elaborada pela autora

Após a definição dos parâmetros para construção dos hologramas, o modelo HQSAR é construído utilizando o método de análise estatística PLS. Durante essa etapa, cada um dos hologramas moleculares será relacionado com a atividade biológica dos compostos do conjunto treinamento. A partir desse modelo, a predição da atividade de novos compostos, semelhantes aos presentes na construção do modelo, poderá ser feita. Além da capacidade de predição, os modelos HQSAR tem como resultado final os chamados mapas de contribuição, que são a representação da estrutura dos compostos em cores que correspondem à influência (positiva/neutra/negativa) de cada átomo na atividade biológica.⁴⁴

1.3.2 Análise Comparativa dos Campos Moleculares (CoMFA)

O método de QSAR 3D da Análise Comparativa dos Campos Moleculares (CoMFA)⁴⁵ é um dos mais utilizados no processo de otimização de compostos líderes e se baseia na hipótese de que as moléculas bioativas produzem seus efeitos por meio de interações 3D, representadas por seus campos estereoquímicos e eletrostáticos. As conformações estruturais e as cargas parciais atômicas das moléculas são cruciais para a criação dos modelos, e devem representar as interações intermoleculares no interior da cavidade de ligação do receptor.

O método de CoMFA combina as técnicas de GRID⁴⁶ e PLS para estabelecer as relações entre os campos de interação molecular (MIFs, do inglês, *Molecular Interaction Fields*) e a atividade biológica. Após o alinhamento estrutural, é criada uma grade reticulada (*grid*) ao redor dos compostos com espaçamento definido pelo usuário (padrão de 2 Å). Uma sonda molecular, ou átomo de prova (carbono sp^3 com carga +1), passa por cada ponto de intersecção do *grid*, contabilizando os MIFs entre a sonda e os átomos de cada composto.⁴¹ A Figura 6 apresenta o processo de construção de um modelo CoMFA. Os MIFs obtidos (variáveis independentes) são relacionados aos valores de atividade biológica (variáveis dependentes) utilizando o método de regressão linear PLS.⁴⁷

Figura 6 - Processo de desenvolvimento de um modelo de QSAR 3D CoMFA

Fonte: Elaborada pela autora

Os valores de MIFs eletrostáticos e estéricos em cada um dos pontos de intersecção do *grid* são calculados a partir das funções de potencial de Coulomb (equação 1; E_c = energia de interação de Coulomb, q_i = carga parcial do átomo i do ligante, q_j = carga do átomo de prova, D = constante dielétrica, r_{ij} = distância entre os átomos i do ligante e o ponto do *grid* j , onde está posicionado o átomo de prova) e potencial de Lennard-Jones (equação 2; E_{vdw} = energia de interação de van der Waals, r_{ij} = distância entre os átomos i do ligante e o ponto do *grid* j , onde está posicionado o átomo de prova; A_{ij} e C_{ij} são constantes que dependem do raio de van der Waals dos átomos), respectivamente. Os potenciais apresentam inclinação muito acentuada nas proximidades da superfície do átomo, o que resulta em valores que tendem a infinito quando os átomos se sobrepõem. Para evitar isso, um valor de corte (*cut-off*) arbitrário é definido para limitar as funções (Figura 7).⁴⁷

$$E_c = \sum_{i=1}^n \frac{q_i q_j}{D r_{ij}} \quad (1)$$

$$E_{vdw} = \sum_{i=1}^n (A_{ij} r_{ij}^{-12} - C_{ij} r_{ij}^{-6}) \quad (2)$$

Figura 7 - Representação dos campos eletrostáticos e estéreos nos estudos de CoMFA, calculados pelos potenciais de Coulomb e Lennard-Jones, respectivamente

Fonte: Adaptada de KUBINYI⁴⁷

A consistência interna do modelo CoMFA pode ser verificada por análises de validação cruzada e a qualidade preditiva para compostos externos à construção do modelo é verificada com a predição dos compostos do conjunto teste. Além da capacidade de prever a atividade biológica de novos compostos, o modelo CoMFA entrega informações sobre as características das regiões moleculares com influência na atividade biológica. Essas informações podem ser visualizadas a partir dos mapas de contorno 3D, que evidenciam importantes regiões de substituição associadas ao aumento da potência, afinidade e seletividade.⁴⁵

1.3.3 Análise Comparativa dos Índices de Similaridade Molecular (CoMSIA)

A Análise Comparativa dos Índices de Similaridades Moleculares (CoMSIA)⁴⁸ é um método de QSAR 3D com princípio similar ao CoMFA, no entanto, o que os diferencia são as funções usadas para calcular os descritores moleculares. No caso do CoMSIA, a função utilizada é uma Gaussiana, que permite a inclusão de campos moleculares adicionais capazes de descrever propriedades físico-químicas relevantes para a atividade biológica. Essa função tem inclinações mais suaves se comparada aos potenciais de Coulomb e Lennard-Jones e não necessita de valores de *cut-off* (Figura 8).

Esse método de QSAR 3D determina as propriedades estereoquímicas, eletrostáticas, hidrofóbicas e de ligação de hidrogênio (doador e aceptor), comumente empregadas, assumindo que esse conjunto de propriedades agrega as contribuições mais importantes responsáveis pelo reconhecimento molecular e a afinidade de um ligante e seu alvo biológico. Os campos moleculares são visualizados no formato de mapas de contorno 3D que indicam os substituintes favoráveis e desfavoráveis para propriedade biológica em estudo.

Figura 8 - Representação da função Gaussiana utilizada no cálculo dos descritores moleculares do modelo CoMSIA

Fonte: Adaptada de KUBINYI⁴⁷

1.4 Planejamento de Novos Antimaláricos

O surgimento de casos de resistência aos fármacos antimaláricos disponíveis representa um grave problema para o controle e tratamento da malária.⁴⁹ Uma estratégia amplamente empregada é a associação de fármacos que agem por diferentes mecanismos de ação, entretanto, essas associações têm gerado cepas multirresistentes que agravam ainda mais o problema de incidência e mortalidade da doença.³ Nesse cenário, torna-se de extrema importância a seleção de novos alvos moleculares e o planejamento de agentes quimioterápicos inovadores contra a malária.

Os métodos computacionais têm sido essenciais no desenvolvimento racional de novos fármacos contra doenças tropicais infecciosas.⁵⁰ Os estudos QSAR têm se mostrado importantes no desenvolvimento de moléculas com atividade biológica contra a malária, dado que esses estudos aceleram o processo de otimização de compostos, aumentando a eficiência

do processo de desenvolvimento de compostos bioativos.³⁴ Sendo assim, os métodos de QSAR são uma ótima ferramenta computacional a ser utilizada na busca de novas estratégias para combater a malária.

1.4.1 N-Miristoiltransferase

A N-miristoilação refere-se à reação bioquímica entre um ácido graxo de 14 carbonos, o miristato - geralmente de uma molécula de miristoil-coenzima A (Figura 9) - e uma glicina N-terminal da proteína alvo.⁵¹ Essa reação irreversível ocorre durante ou após a etapa de tradução na célula e é catalisada pela enzima N-miristoiltransferase (NMT), enzima essencial e exclusiva de eucariotos. Proteínas alvo da N-miristoilação apresentam diferentes funções intracelulares. Muitas dessas proteínas são componentes de cascatas de transdução de sinal e, devido à ligação com o miristato, são localizadas em membranas celulares.⁵²

Nesse contexto, a proteína NMT (EC 2.3.1.97 – Figura 10) tornou-se um alvo terapêutico atrativo dada a importância dos processos biológicos envolvidos, presentes em diversos patógenos humanos.⁵³ Recentemente, a NMT foi validada como alvo biológico em *Plasmodium*. A inibição seletiva da enzima de *P. falciparum* causa danos irreversíveis na montagem do complexo de membrana celular interna levando o parasita a morte.⁵⁴

Figura 9 - Estrutura molecular da miristoil-coenzima A

Fonte: Elaborada pela autora

Figura 10 - Estruturas 3D da NMT humana (código PDB: 4C2Y). (A) Modelo de fitas. (B) Modelo de superfície com destaque para a cavidade de ligação da molécula de miristoil-CoA

Fonte: Elaborada pela autora

1.4.2 Compostos derivados de benzotiofeno

Pesquisas realizadas previamente para candidatos a fármacos antimaláricos resultaram na identificação dos derivados 1-benzotiofenos (Figura 11) com atividade inibitória frente a NMT de *P. falciparum*.^{55,56} Uma série de derivados 1-benzotiofenos foi sintetizada e avaliada contra as NMTs de diferentes organismos (e.g., *Plasmodium vivax*, *Plasmodium falciparum* e *Homo sapiens*).^{57,58} Os derivados 1-benzotiofenos mostraram-se relevantes para o estudo de inibição e seletividade da NMT. Esses estudos forneceram as bases para a organização de dois conjuntos de dados com elevado potencial para o desenvolvimento de estudos de QSAR para a NMT de *P. falciparum* e para a NMT humana.

Figura 11 - Estrutura molecular do 1-benzotiofeno

Fonte: Elaborada pela autora

OBJETIVOS

2 OBJETIVOS

O objetivo fundamental desse estudo é estabelecer as bases estruturais e moleculares para a atividade biológica de uma série de inibidores da enzima NMT de *P. falciparum* como candidatos a fármacos antimaláricos. Para alcançar esse objetivo fundamental foram estabelecidos os seguintes objetivos específicos:

- Construção e padronização de conjuntos de dados de moléculas bioativas derivadas do 1-benzotiofeno com atividade inibitória frente às enzimas NMT de *P. falciparum* e de *H. sapiens*;
- Construção de modelo estrutural da enzima *Pf*NMT para guiar os estudos de modelagem molecular;
- Estudo das relações quantitativas entre estrutura e atividade pela aplicação de métodos de QSAR 2D e QSAR 3D para os conjuntos de dados da *Pf*NMT e da *Hs*NMT;
- Identificação das bases estruturais envolvidas no processo de reconhecimento molecular para o desenvolvimento de novos compostos candidatos a fármacos antimaláricos.

ESTRATÉGIAS E MÉTODOS

3 ESTRATÉGIAS E MÉTODOS

3.1 Estratégias computacionais

Todos os trabalhos de planejamento baseado na estrutura do ligante (LBDD), construções de modelos QSAR 2D e 3D, e também a construção do modelo estrutural por homologia foram realizados em estações computacionais do CIBFar-CEPID. Estas estações operam no sistema operacional Linux, utilizando-se, no desenvolvimento desse trabalho, os módulos QSAR da interface plataforma SYBYL-X 2.0 (Tripos, St. Louis, EUA).

3.2 Criação de conjuntos de dados para modelagem de QSAR

Os conjuntos de dados coletados consistem em 39 estruturas, sendo 37 derivadas de 1-benzotiofeno e 2 derivadas do 1-benzofurano, com a medida experimental da propriedade biológica correspondente (e.g., valores de K_i) contra as enzimas *Pf*NMT e *Hs*NMT. A atividade biológica K_i foi convertida para $pK_i = -\log K_i$ para ser utilizada como variável dependente nos modelos QSAR. As estruturas dos compostos foram desenhadas e, em seguida, exportadas para o programa SYBYL-X onde tiveram a estrutura molecular otimizada para obtenção da conformação de menor energia, utilizando o campo de força Tripos⁵⁹ com o algoritmo de convergência conjugado Powell⁶⁰, com critério de parada o gradiente da força RMS (raiz do desvio quadrático médio, do inglês, *Root Mean Square*) entre iterações menor que 0,005 kcal/mol.Å. Os compostos foram divididos em conjunto treinamento e conjunto teste na proporção 5:1 para posterior construção e validação interna e externa dos modelos, de forma que ambos conjuntos fossem representativos em estrutura molecular e atividade biológica do conjunto total. A Tabela 3 representa a estrutura molecular de todos os compostos do conjunto de dados e respectivas atividades biológicas.

Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens*

Composto	Estrutura química	pK_i	
		<i>Pf</i> NMT	<i>Hs</i> NMT
1		4,87	4,00
2		5,35	4,00
3		5,89	4,00
4		5,47	5,36
5		6,08	5,32

continua

continuação

Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens* (continuação)

Composto	Estrutura química	p <i>K_i</i>	
		<i>Pf</i> NMT	<i>Hs</i> NMT
6		4,97	4,00
7		4,00	5,50
8		6,17	4,98
9		6,92	5,66
10		7,59	7,51

continua

continuação
Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens* (continuação)

Composto	Estrutura química	p <i>K_i</i>	
		<i>Pf</i> NMT	<i>Hs</i> NMT
11		5,40	5,94
12		5,95	5,73
13		5,00	4,00
14		5,00	4,00
15		5,00	4,00
16		7,15	6,10

continua

continuação

Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens* (continuação)

Composto	Estrutura química	p <i>K_i</i>	
		<i>Pf</i> NMT	<i>Hs</i> NMT
17		6,61	5,81
18		7,29	6,72
19		6,79	6,10
20		7,82	6,68
21		6,92	5,93
22		7,30	7,00

continua

continuação
Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens* (continuação)

Composto	Estrutura química	p <i>K_i</i>	
		<i>Pf</i> NMT	<i>Hs</i> NMT
23		7,82	7,30
24		6,57	6,07
25		7,11	5,86
26		7,48	6,37
27		6,48	5,68
28		6,82	5,81

continua

continuação

Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens* (continuação)

Composto	Estrutura química	p <i>K_i</i>	
		<i>Pf</i> NMT	<i>Hs</i> NMT
29		7,01	6,01
30		6,41	5,56
31		6,33	5,72
32		6,67	5,40
33		6,90	5,61
34		7,26	5,81

continua

conclusão
Tabela 3 - Estruturas químicas dos derivados 1-benzotiofenos com atividade *in vitro* frente à NMT de *P. falciparum* e *H. sapiens* (continuação)

Composto	Estrutura química	p <i>K_i</i>	
		<i>Pf</i> NMT	<i>Hs</i> NMT
35		6,95	5,98
36		8,10	7,22
37		6,65	5,65
38		6,73	6,15
39		6,91	6,08

Fonte: Elaborada pela autora

3.3 Modelagem QSAR: QSAR 2D e QSAR 3D

Os estudos de QSAR 2D e QSAR 3D visam à geração de modelos preditivos para conjuntos de moléculas bioativas.^{34,61-65} Foram desenvolvidos modelos de QSAR 2D e 3D utilizando o programa SYBYL-X para os dois conjuntos de dados, da *Pf*NMT e da *Hs*NMT, e ao final os resultados foram comparados, buscando informações relevantes sobre seletividade para a enzima do parasita.

3.3.1 QSAR 2D – Holograma QSAR

Os modelos de QSAR 2D foram gerados utilizando o módulo HQSAR do programa SYBYL_X. Após a divisão dos conjuntos treinamento e teste, os parâmetros de geração dos hologramas moleculares foram variados buscando o modelo mais preditivo. Inicialmente, foram gerados modelos com diferentes combinações de distinção de fragmento - átomos (A), ligações (B), conectividade (C), presença de átomos de hidrogênio (H), quiralidade (Ch) e doadores e aceptores de ligação de hidrogênio (DA) – mantendo o parâmetro tamanho de fragmentos 4-7. A busca por modelos com maior consistência interna variou, para cada configuração de distinção de fragmento, entre os comprimentos de holograma padrões (53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, 401) e também comprimentos de holograma maiores (499, 673, 809, 997). A influência do tamanho dos fragmentos também foi avaliada, calculando modelos com tamanhos: 2-5, 3-6, 4-7, 5-8, 6-9 e 7-10 átomos.

Os modelos HQSAR foram construídos pela correlação dos hologramas moleculares e propriedades biológicas pK_i , utilizando a análise PLS⁶⁶ (do inglês *Partial Least Square*). Diferentes combinações dos parâmetros de distinção de fragmento, tamanho de fragmento e comprimento de holograma foram analisadas em busca do melhor modelo. A validação interna foi realizada usando o método de validação cruzada *leave-one-out* (LOO). Os parâmetros estatísticos analisados foram os coeficientes de correlação convencional r^2 (equação 3; Y_a , valor experimental de atividade biológica, Y_c , valor de atividade biológica calculada pelo modelo, Y_m , valor médio de atividade biológica de todas as moléculas do conjunto treinamento) e de validação cruzada LOO q^2 (equação 4; Y_p , valor predito de atividade biológica), para os quais se espera valores $r^2 > 0,6$ e $q^2 > 0,5$.^{41,67} A validação externa foi desenvolvida utilizando o conjunto teste previamente selecionado. Os compostos desse conjunto tiveram sua atividade biológica predita e a capacidade de predição dos modelos foi avaliada pelo coeficiente de correlação de predição r^2_{pred} , calculado pela equação 4.⁶⁷

$$r^2 = 1 - \frac{\sum_{i=1}^N (Y_c - Y_a)^2}{\sum_{i=1}^N (Y_m - Y_a)^2} \quad (3)$$

$$q^2 = 1 - \frac{PRESS}{SD} = 1 - \frac{\sum_{i=1}^N (Y_a - Y_p)^2}{\sum_{i=1}^N (Y_a - Y_m)^2} \quad (4)$$

3.3.2 QSAR 3D – CoMFA e CoMSIA

3.3.2.1 Cargas parciais atômicas

Os cálculos de cargas parciais atômicas e de conformação estrutural, bem como o alinhamento dos compostos, são aspectos determinantes na construção de modelos QSAR 3D. Os cálculos de cargas podem ser mais rápidos e com aproximações, como no caso das cargas empíricas e semi-empíricas, ou com maior custo computacional e mais precisos, caso dos métodos quânticos.

Os cálculos das cargas parciais empíricas, Gasteiger-Hückel⁶⁸ e MMFF94⁶⁹, foram realizados no programa SYBYL-X, o método semi-empírico AM1-BCC⁷⁰ foi desenvolvido no *software* AmberTools 16 utilizando a ferramenta *antechamber*⁷¹⁻⁷², os métodos quânticos foram realizados pelo programa ORCA⁷³, que calculou as cargas CHELPG⁷⁴ e Mulliken Population Analysis (MPA)⁷⁵, utilizando o conjunto de bases 6-31G*⁷⁶⁻⁷⁷ e o método de densidade funcional B3LYP⁷⁸. A construção desses diferentes modelos visa a identificação do melhor método para a descrição/correlação da atividade inibitória dos derivados benzotiofênicos.⁷⁹⁻⁸¹ Os cálculos das cargas foram realizados considerando a geometria otimizada de acordo com cada nível de teoria dos métodos. Essas conformações foram assumidas como a conformação bioativa dos compostos.

3.3.2.2 Alinhamento molecular

O alinhamento estrutural dos compostos foi feito utilizando dois métodos baseados na estrutura do ligante e um baseado na estrutura do alvo molecular, utilizando dois programas de docagem molecular. Um dos métodos de alinhamento baseado na estrutura do ligante foi feito a partir da máxima subestrutura comum (MCS, do inglês, *Maximum Common Substructure*) aplicada pela função *align database* do programa SYBYL-X, considerando como estrutura de

referência o composto de maior atividade biológica, **36**. Dessa forma, todos os compostos com estruturas previamente otimizadas foram sobrepostos com base no menor desvio quadrático médio (RMSD, do inglês *Root Mean Square Deviation*) entre os átomos da subestrutura selecionada, sem alterar as conformações.

O segundo método de alinhamento estrutural baseado na estrutura do ligante foi o alinhamento flexível de ligantes (ferramenta *Flexible Ligand Alignment*) desenvolvido pelo programa Schrödinger. Essa ferramenta faz uma busca conformacional com o método ConfGen⁸² e alinha os confôrmers a partir de uma molécula referência. O confômero de cada ligante com a sobreposição mais aproximada da molécula referência é selecionado. Neste estudo, a molécula utilizada como referência foi a **34** cristalográfica obtida da estrutura PDB 4CAF.

Além do alinhamento baseado na estrutura do ligante, foi desenvolvido o alinhamento baseado na estrutura do alvo molecular. A estrutura da *Pf*NMT, construída neste trabalho pelo método de modelagem por homologia, foi utilizada para a docagem dos compostos. As docagens moleculares foram realizadas no SYBYL-X com o programa Surflex-Dock⁸³ e no Schrödinger com o programa Glide.⁸⁴

Na docagem pelo Surflex-Dock, os átomos de hidrogênio foram adicionados na geometria padrão à estrutura do alvo molecular pelo módulo BIOPOLYMER do SYBYL-X. O sítio ativo foi definido a partir da análise da estrutura experimentalmente elucidada da NMT de *P. vivax* (*Pv*NMT, PDB: 4CAF) cocrystalizada com um dos compostos benzotiofênicos (composto **34**). Os resíduos correspondentes do sítio ativo da *Pf*NMT foram selecionados considerando um raio de 8 Å a partir da posição do ligante cristalográfico na estrutura de *Pv*NMT. Durante a docagem, foi permitida a flexibilidade dos hidrogênios e átomos pesados da proteína e o composto **34** na conformação cristalográfica foi utilizado como estrutura referência. Para cada um dos compostos, foram geradas dez poses, mas somente as conformações melhores pontuadas foram consideradas para a construção dos modelos QSAR 3D.

Na docagem pelo Glide, a estrutura do alvo molecular foi preparada com parâmetros padrão pelo módulo *Protein Preparation Wizard* do programa Schrödinger, que faz a adição dos hidrogênios e minimização da estrutura. Os ligantes foram preparados a partir do módulo *LigPrep* com parâmetros padrão. A definição do sítio de ligação foi realizada pelo módulo *Receptor Grid Generation*, selecionando uma caixa de lado 10 Å com centroide calculado a partir dos resíduos de aminoácido do sítio de ligação dos benzotiofenos. Foram geradas 10

poses para cada um dos compostos e a pose com melhor pontuação foi selecionada para fazer parte do conjunto de dados dos modelos QSAR 3D.

3.3.2.3 Construção dos modelos

Os modelos CoMFA e CoMSIA foram construídos com base na energia de interação molecular entre os átomos dos compostos posicionados no interior de uma grade tridimensional reticulada e a sonda molecular apropriada que passa em cada uma das intersecções dos retículos da grade. As energias de interação calculadas em cada intersecção correspondem a uma tabela contendo as colunas de descritores moleculares (variáveis independentes) que serão correlacionados com a propriedade biológica (variável dependente) a partir do método de regressão SAMPLS (do inglês, *SAM*ple-*distance* PLS), que é uma modificação mais rápida da análise PLS.

A grade tridimensional apresenta como padrão a distância de 2 Å entre as intersecções. Um carbono sp^3 com carga +1 foi utilizado para calcular os campos do CoMFA de acordo com os parâmetros padrão do SYBYL para campos estéreo e eletrostáticos. Para o CoMSIA, foram calculados os campos estéreo, eletrostáticos, hidrofóbicos e de ligação de hidrogênio, sendo a sonda molecular padrão com carga +1, raio atômico 1, hidrofóbica, com propriedade de ligação de hidrogênio +1 e fator de atenuação da função Gaussiana 0,3. Assim como para o método de HQSAR, a validação interna dos modelos foi realizada usando o método de LOO. Os parâmetros estatísticos analisados foram os coeficientes r^2 e q^2 (LOO) e o r^2_{pred} avaliou a capacidade de predição dos modelos.

Os modelos CoMFA podem ser otimizados utilizando o procedimento de focagem de região (do inglês, *region focusing*) implementado pelo próprio método PLS do CoMFA. Como padrão, o cálculo dos descritores moleculares no método CoMFA considera o mesmo peso para os valores de potencial calculados em cada ponto de intersecção do *grid*. O procedimento de focagem de região gera um refinamento do modelo por atribuir maiores pesos para os pontos de intersecção mais relevantes para a atividade biológica.

3.4 Modelagem por homologia

A estrutura molecular da enzima PfNMT foi construída a partir do método de modelagem por homologia utilizando o programa Modeller 9.10⁸⁵. A estrutura cristalográfica de alta resolução da proteína PvNMT foi utilizada como molde estrutural (código PDB: 4CAF).

A sequência completa de aminoácidos da estrutura-alvo *Pf*NMT foi obtida no formato FASTA pela base de dados de sequências de aminoácidos NCBI (GI: 6581106).

O alinhamento das estruturas primárias da NMT de *P. falciparum* e *P. vivax* foi realizado utilizando o programa ClustalW2. Após o alinhamento das sequências, 10 modelos foram gerados automaticamente usando os parâmetros padrão do Modeller. A função de energia potencial DOPE⁸⁶ (do inglês, *Discrete Optimized Potential Energy*) foi utilizada para selecionar o modelo com menor energia global. A etapa de validação incluiu a verificação do diagrama de Ramachandran⁸⁷ e análise estrutural pelo programa PROCHECK⁸⁸ e servidor Molprobity⁸⁹. O valor de RMSD entre os átomos das cadeias principais das estruturas molde e alvo foi calculado pela sobreposição estrutural no programa PyMOL (The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC).

RESULTADOS E DISCUSSÕES

4 RESULTADOS E DISCUSSÕES

4.1 Preparo dos conjuntos de dados

Os conjuntos de dados foram cuidadosamente separados em conjunto treinamento (31 compostos) e conjunto teste (8 compostos) de forma a serem representativos do conjunto total, considerando as estruturas moleculares e atividades biológicas. Para tanto, foi analisada a distribuição do valor de pK_i dos compostos e a diversidade estrutural, com ajuda da ferramenta do SYBYL-X que gera mapas de similaridade estrutural baseado no descritor *UNITY Fingerprint*. A Figura 12 apresenta a distribuição do pK_i em cada um dos conjuntos.

Figura 12 - Representação gráfica da distribuição de pK_i para os conjuntos treinamento, teste e total, mostrando a representatividade dos conjuntos de dados da *PfNMT* (A) e *HsNMT* (B)

Fonte: Elaborada pela autora

4.2 Construção dos modelos HQSAR

4.2.1 Modelos HQSAR para o conjunto de dados da *PfNMT*

A construção dos modelos de HQSAR iniciais foi realizada para várias combinações de distinção de fragmento, mantendo o tamanho padrão dos fragmentos (4-7). Os comprimentos de holograma considerados para a construção dos modelos foram 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, 401. Os resultados estatísticos obtidos a partir da análise PLS dos modelos iniciais para o conjunto de dados da *PfNMT* estão apresentados na Tabela 4.

Tabela 4 - Resultados estatísticos dos modelos HQSAR iniciais para os inibidores benzotiofênicos contra a enzima P/NMT, com tamanho de fragmento de 4 – 7 átomos

Modelo	Distinção de Fragmento	q^2	r^2	SEE	HL	N
1	A/B	0,73	0,90	0,32	401	3
2	A/B/C	0,75	0,92	0,30	401	3
3	A/B/C/H	0,71	0,94	0,26	353	5
4	A/B/C/H/Ch	0,69	0,94	0,25	353	5
5	A/B/C/H/DA	0,71	0,91	0,32	307	3
6	A/B/C/H/Ch/DA	0,70	0,93	0,28	307	4
7	A/B/C/DA	0,74	0,99	0,33	53	4
8	A/B/Ch/DA	0,73	0,92	0,29	401	4
9	A/B/H/DA	0,66	0,92	0,30	307	4
10	A/C/H/Ch/DA	0,64	0,89	0,34	83	4
11	A/C/DA	0,73	0,88	0,36	71	3
12	B/C/DA	0,73	0,93	0,28	83	5
13	B/C/H/DA	0,73	0,93	0,28	83	5
14	B/H/DA	0,70	0,91	0,31	257	4

Distinção de fragmento: A – átomo, B – ligação, C – conectividade, H – hidrogênio, Ch – quiralidade, DA – doadores e aceptores de ligação de hidrogênio; q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento de holograma; N, número ótimo de componentes principais

Fonte: Elaborada pela autora

Os modelos iniciais resultaram em parâmetros estatísticos q^2 e r^2 satisfatórios de acordo com o critério exigido ($q^2 > 0,5$ e $r^2 > 0,6$). No entanto, os valores de erro padrão de correlação apresentaram-se mais altos do que esperado (critério: $SEE < 0,3$), o que poderia indicar a presença de compostos *outliers* no conjunto treinamento. A fim de encontrar possíveis *outliers*, cada um dos compostos foi excluído do conjunto treinamento e os parâmetros estatísticos dos novos modelos construídos foram analisados. Com a exclusão do composto 7, os parâmetros estatísticos apresentaram-se ainda melhores e os valores de SEE diminuíram sensivelmente. Esses resultados estatísticos obtidos para o novo conjunto treinamento (30 compostos) estão apresentados na Tabela 5.

Tabela 5 - Resultados estatísticos dos modelos HQSAR para os inibidores benzotiofênicos contra a enzima PfNMT com a retirada do composto *outlier* 7 e tamanho de fragmento de 4 – 7 átomos

Modelo	Distinção de Fragmento	q^2	r^2	SEE	HL	N
1	A/B	0,77	0,92	0,26	401	3
2	A/B/C	0,80	0,96	0,20	401	4
3	A/B/C/H	0,81	0,99	0,12	353	6
4	A/B/C/H/Ch	0,78	0,99	0,12	353	6
5	A/B/C/H/DA	0,83	0,98	0,15	307	5
6	A/B/C/H/Ch/DA	0,85	0,99	0,08	307	6
7	A/B/C/DA	0,79	0,97	0,17	53	6
8	A/B/Ch/DA	0,79	0,97	0,17	401	6
9	A/B/H/DA	0,78	0,98	0,15	401	5
10	A/C/H/Ch/DA	0,79	0,98	0,15	59	6
11	A/C/DA	0,79	0,91	0,28	71	3
12	B/C/DA	0,80	0,96	0,20	83	5
13	B/C/H/DA	0,80	0,96	0,20	83	5
14	B/H/DA	0,73	0,93	0,25	353	4

Distinção de fragmento: A – átomo, B – ligação, C – conectividade, H – hidrogênio, Ch – quiralidade, DA – doadores e aceptores de ligação de hidrogênio; q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento de holograma; N, número ótimo de componentes principais

Fonte: Elaborada pela autora

O modelo 5 (A/B/C/H/DA) apresentou os melhores parâmetros estatísticos ($q^2 = 0,83$, $r^2 = 0,96$ e $SEE = 0,15$) e foi selecionado para a próxima etapa de análise da influência do tamanho dos fragmentos sobre a qualidade dos modelos. Nessa etapa, foram considerados os tamanhos de fragmento 2-5, 3-6, 5-8, 6-9 e 7-10. Além disso, analisou-se a influência do aumento do comprimento de holograma nos modelos HQSAR, incluindo os comprimentos 499, 673, 809 e 997. Os resultados dessa etapa estão apresentados na Tabela 6.

Os dados da Tabela 6 indicam que o valor mais alto de q^2 foi obtido para o tamanho de fragmento 3-6. Apesar disso, esse modelo apresentou maior número de componentes principais, o que pode ter como consequência modelos com sobreajuste dos dados. Os maiores comprimentos de holograma não resultaram em melhores modelos. Por fim, uma última etapa foi realizada para avaliar o comportamento do modelo selecionado anteriormente com a diminuição do N.

Tabela 6 - Resultados estatísticos dos modelos HQSAR dos inibidores benzotiofênicos contra a enzima *Pf*NMT, variando o tamanho dos fragmentos e mantendo a distinção de fragmento A/B/C/H/DA

Tamanho do fragmento	q^2	r^2	SEE	HL	N
2 – 5	0,76	0,91	0,27	199	3
3 – 6	0,85	0,98	0,15	61	6
4 – 7	0,83	0,98	0,15	307	5
5 – 8	0,82	0,98	0,12	151	5
6 – 9	0,83	0,99	0,08	257	6
7 – 10	0,80	0,99	0,08	151	6

q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento de holograma; N, número ótimo de componentes principais

Fonte: Elaborada pela autora

Com a diminuição de N, o modelo não teve perda significativa na consistência interna e, portanto, o modelo com N = 3 foi selecionado como modelo final para etapa de validação externa. O modelo HQSAR final construído com o conjunto de dados da *Pf*NMT tem como parâmetros a distinção de fragmento A/B/C/H/DA, tamanho de fragmento 4-7, comprimento de holograma = 307, N = 3, $r^2 = 0,95$, $q^2 = 0,81$. Os resultados da etapa de validação externa estão apresentados na Tabela 7.

Tabela 7 - Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo HQSAR da *Pf*NMT

Composto	pK_i		
	Experimental	Predito	Resíduo*
5	6,08	6,46	0,38
12	5,95	5,69	-0,26
14	5,00	5,04	0,04
16	7,15	6,77	-0,38
23	7,82	7,99	0,17
26	7,48	7,46	-0,02
35	6,95	7,29	0,34
37	6,65	5,85	-0,80

*Resíduo = pK_i experimental – pK_i predito

Fonte: Elaborada pela autora

Com a finalidade de avaliar a capacidade preditiva do modelo, além da consistência interna e da correlação dos dados do conjunto treinamento, deve-se analisar as predições de

propriedade biológica para compostos não utilizados na construção do modelo. Essa etapa de validação externa faz a predição de pK_i dos compostos do conjunto teste e, a partir dos valores preditos, calcula-se o parâmetro r^2_{pred} .

O r^2_{pred} é um parâmetro que estima a capacidade preditiva de um modelo considerando ambos conjuntos, treinamento e teste. Empregando esse cálculo, obteve-se para o melhor modelo um $r^2_{pred} = 0,81$, indicando que este modelo possui boa capacidade de predição para novas moléculas. A representação gráfica de pK_i predito *versus* pK_i experimental ilustra a concordância entre esses valores (Figura 13).

Em complemento à predição da propriedade biológica, o HQSAR fornece informações sobre a relação entre estrutura e atividade em forma de mapas de contribuição. Esses mapas representam a influência de cada átomo dos compostos na atividade biológica. Átomos representados em vermelho, vermelho-alaranjado e laranja indicam contribuição negativa, átomos em amarelo, amarelo-esverdeado e verde indicam contribuição positiva e átomos em branco, contribuição neutra. A Figura 14 exhibe os mapas de contribuição dos compostos benzotiofênicos de maior e menor atividade biológica contra a enzima *PfNMT*, obtidos a partir do melhor modelo de HQSAR construído.

O mapa de contribuição do composto de menor atividade apresenta no anel benzofurano uma região de contribuição negativa colorida em laranja, indicando que esse grupo é desfavorável para a atividade inibitória. No caso do mapa de contribuição do composto de maior atividade, várias regiões apresentam-se favoráveis à atividade biológica, com destaque para o átomo de enxofre do anel benzotiofeno, o anel oxadiazol e o grupo metileno. No geral, os mapas mostram a influência favorável do anel oxadiazol ligados ao anel benzotiofeno em contrapartida aos grupos amida nessa mesma região. Como exemplo, os compostos **13** e **15** que apresentam em seus mapas de contribuição o grupo amida em vermelho e, conseqüentemente, tem baixa atividade inibitória ($pK_i = 5$).

Figura 13 - Representação gráfica dos valores experimentais *versus* preditos de pK_i para o melhor modelo HQSAR do conjunto de dados da *PfNMT*, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos

Fonte: Elaborada pela autora

Figura 14 - Mapas de contribuição dos compostos de menor atividade biológica, à esquerda, e de maior atividade, à direita, obtidos a partir do melhor modelo HQSAR construído com o conjunto de dados da *PfNMT*

Fonte: Elaborada pela autora

4.2.2 Modelo HQSAR para o conjunto de dados da *HsNMT*

As mesmas etapas seguidas na construção dos modelos HQSAR para o conjunto de dados da *PfNMT* foram aplicadas para o conjunto de dados da *HsNMT*. Os modelos iniciais mostraram valores aceitáveis de q^2 e r^2 . No entanto, o erro padrão da correlação observado foi

maior do que esperado, o que poderia indicar a presença de compostos *outlier* no conjunto treinamento (Tabela 8).

Tabela 8 - Resultados estatísticos dos modelos HQSAR iniciais para os inibidores benzotiofênicos contra a enzima HsNMT, com tamanho de fragmento de 4 – 7 átomos

Modelo	Distinção de Fragmento	q^2	r^2	SEE	HL	N
1	A/B	0,58	0,77	0,47	59	2
2	A/B/C	0,55	0,84	0,40	353	3
3	A/B/C/H	0,45	0,85	0,40	83	4
4	A/B/C/H/Ch/	0,43	0,78	0,47	83	3
5	A/B/C/H/Ch/DA	0,50	0,82	0,43	257	3
6	A/B/C/H/DA	0,52	0,82	0,43	257	3
7	A/B/C/Ch/DA	0,55	0,83	0,42	401	3
8	A/B/H/Ch/DA	0,52	0,82	0,43	199	3
9	A/B/C/DA	0,57	0,82	0,42	257	3
10	A/B/Ch/DA	0,53	0,81	0,44	353	3
11	A/C/H/DA	0,54	0,83	0,41	199	3
12	A/B/H/DA	0,49	0,81	0,44	257	3
13	A/B/DA	0,54	0,85	0,39	199	4
14	B/C/DA	0,56	0,83	0,41	401	3

Distinção de fragmento: A – átomo, B – ligação, C – conectividade, H – hidrogênio, Ch – quiralidade, DA – doadores e aceptores de ligação de hidrogênio; q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento de holograma; N, número ótimo de componentes principais

Fonte: Elaborada pela autora

Os modelos foram reconstruídos excluindo cada um dos compostos do conjunto de dados por vez para análise dos parâmetros estatísticos. Com isso, duas moléculas foram identificadas como *outliers* (compostos **6** e **10**) e excluídas do conjunto treinamento. Pela diminuição da quantidade de compostos, foi necessário retirar um composto do conjunto teste e adicioná-lo ao conjunto treinamento. Os resultados estatísticos obtidos com novo conjunto treinamento (30 compostos) estão exibidos na Tabela 9. Para a construção desses modelos, também foram incluídos os maiores tamanhos de holograma.

Tabela 9 - Resultados estatísticos dos modelos HQSAR para os inibidores benzotiofênicos contra a enzima HsNMT com a retirada dos compostos *outliers* 6 e 10 e tamanho de fragmento de 4 – 7

Modelo	Distinção de Fragmento	q^2	r^2	SEE	HL	N
1	A/B	0,81	0,93	0,26	809	3
2	A/B/C	0,78	0,95	0,23	353	4
3	A/B/C/H	0,68	0,96	0,21	257	5
4	A/B/C/H/Ch/	0,67	0,96	0,21	257	5
5	A/B/C/H/Ch/DA	0,73	0,96	0,19	61	5
6	A/B/C/H/DA	0,72	0,91	0,29	257	3
7	A/B/C/Ch/DA	0,76	0,94	0,23	673	4
8	A/B/H/Ch/DA	0,72	0,96	0,21	997	5
9	A/B/C/DA	0,79	0,93	0,26	673	3
10	A/B/Ch/DA	0,78	0,94	0,24	809	4
11	A/C/H/DA	0,72	0,93	0,25	809	3
12	A/B/H/DA	0,70	0,94	0,25	997	4
13	A/B/DA	0,78	0,94	0,25	809	3
14	B/C/DA	0,80	0,93	0,26	673	3

Distinção de fragmento: A – átomo, B – ligação, C – conectividade, H – hidrogênio, Ch – quiralidade, DA – doadores e aceptores de ligação de hidrogênio; q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento de holograma; N, número ótimo de componentes principais

Fonte: Elaborada pela autora

Os melhores modelos, 1, 9 e 14, foram selecionados para as etapas seguintes. O aumento do tamanho dos hologramas resultou em melhores resultados estatísticos. Por fim, como melhor modelo foi escolhido o 9 que apresentou melhor valor de r^2_{pred} . A variação do tamanho de fragmento foi investigada para analisar a influência sobre a qualidade dos parâmetros estatísticos (Tabela 10).

A variação do tamanho de fragmento resultou em parâmetro q^2 melhor para o tamanho de fragmento 6-9. Em contrapartida, esse modelo exibe um número de componentes principais mais alto (N=6), o que pode indicar sobreajuste dos dados.

Tabela 10 - Resultados estatísticos dos modelos HQSAR dos inibidores benzotiofênicos contra a enzima *HsNMT*, variando o tamanho dos fragmentos e mantendo a distinção de fragmento A/B/C/DA

Tamanho do fragmento	q^2	r^2	SEE	HL	N
2 – 5	0,75	0,92	0,27	199	4
3 – 6	0,78	0,94	0,24	673	4
4 – 7	0,79	0,93	0,26	673	3
5 – 8	0,77	0,95	0,22	673	5
6 – 9	0,81	0,96	0,21	61	6
7 – 10	0,77	0,94	0,24	61	5

q^2 , coeficiente de correlação da validação cruzada; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; HL, comprimento de holograma; N, número ótimo de componentes principais

Fonte: Elaborada pela autora

O modelo HQSAR final construído com o conjunto de dados da *HsNMT* tem como parâmetros a distinção de fragmento A/B/C/DA, tamanho de fragmento 4-7, comprimento de holograma = 673, N = 3, $r^2 = 0,93$, $q^2 = 0,79$. Os resultados da etapa de validação externa estão apresentados na Tabela 11.

Obteve-se para o melhor modelo um $r^2_{pred} = 0,64$, indicando que este modelo possui capacidade de predição satisfatória para novas moléculas. A concordância entre esses valores preditos e experimentais pode ser observada na representação gráfica dos valores de pK_i predito versus pK_i experimental (Figura 15).

Os mapas de contribuição gerados a partir do melhor modelo HQSAR para o conjunto de dados da *HsNMT* forneceram informações importantes sobre as estruturas e suas atividades biológicas. A Figura 16 exibe a comparação dos mapas de contribuição obtidos para os modelos da *HsNMT* e da *PfNMT* para os compostos de maior atividade inibitória e de maior diferença de pK_i .

Comparando-se os mapas de contribuição da *PfNMT* e da *HsNMT* para cada um dos compostos, é possível notar várias regiões com influência positiva na atividade inibitória em comum para as duas enzimas. A fim de melhorar a seletividade dos inibidores para a enzima do parasita, deve-se explorar regiões que influenciem positivamente apenas para a *PfNMT* e/ou negativamente para *HsNMT*. Regiões como o anel benzeno do benzotiofeno poderiam ser modificados na tentativa de diminuir a atividade contra a enzima humana, dado que apenas nos mapas de contribuição da *HsNMT* essa região indica influência positiva.

Tabela 11 - Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo HQSAR da *HsNMT*

Composto	pK_i		
	Experimental	Predito	Resíduo*
3	4,00	5,38	1,38
12	5,73	5,34	-0,39
19	6,06	5,96	-0,10
23	7,30	6,99	-0,32
26	6,37	6,07	-0,30
29	6,01	5,80	-0,21
34	5,81	6,20	0,39

*Resíduo = pK_i experimental – pK_i predito

Fonte: Elaborada pela autora

Figura 15 - Representação gráfica dos valores experimentais *versus* preditos de pK_i para o melhor modelo HQSAR do conjunto de dados da *HsNMT*, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos

Fonte: Elaborada pela autora

Figura 16 - Mapas de contribuição dos compostos de maior atividade biológica (composto **36**) e de maior diferença de pK_i (composto **33**) entre as enzimas *PfNMT* e *HsNMT* obtidos a partir dos melhores modelos HQSAR

Fonte: Elaborada pela autora

4.3 Construção do modelo estrutural da enzima *PfNMT*

A falta de uma estrutura experimentalmente elucidada da enzima *PfNMT* apontou a necessidade de construir um modelo computacional para a estrutura tridimensional dessa enzima. O modelo foi posteriormente utilizado na etapa de alinhamento molecular do método de QSAR 3D. O alinhamento das sequências de aminoácidos da proteína alvo (*PfNMT*) e proteína molde (*PvNMT*), realizado utilizando o programa ClustalW2, indicou identidade sequencial de 81% (Figura 17). A identidade entre os resíduos da cavidade de ligação é de 97%, havendo apenas uma mutação, do resíduo Tyr334 na enzima *PvNMT* para Phe334 na *PfNMT*.

Figura 17 - Alinhamento das sequências de aminoácidos da NMT de *P. falciparum* e *P. vivax* com destaque, em azul, para os resíduos idênticos entre as sequências

Fonte: Elaborada pela autora

Entre os dez modelos construídos pelo programa Modeller, foi selecionado como melhor o modelo com menor energia DOPE. A qualidade do modelo tridimensional de *Pf*NMT construído foi avaliada pelo diagrama de Ramachandran, PROCHECK e Molprobit. A Figura 18 indica o resultado do diagrama de Ramachandran e a Figura 19 representa o alinhamento molecular da estrutura molde (*Pv*NMT, PDB: 4CAF) e o modelo construído da *Pf*NMT.

Figura 18 - Diagrama de Ramachandran do modelo tridimensional construído da enzima *Pf*NMT

Fonte: Elaborada pela autora

Figura 19 - Sobreposição da estrutura molde da NMT de *P. vivax* (PDB: 4CAF), em azul, e o modelo construído da *Pf*NMT, em verde. (A) Estruturas totais e (B) Sítio de ligação dos compostos benzotiofênicos após modificação da conformação dos resíduos marcados, com o composto 34 cristalográfico representado em magenta

Fonte: Elaborada pela autora

O modelo apresentou 91% dos resíduos de aminoácidos em regiões favorecidas do diagrama de Ramachandran, 9% em regiões permitidas e nenhum resíduo nas regiões não-permitidas. A análise no servidor Molprobity não indicou problemas relevantes na estrutura. O modelo não apresentou alças com alta energia e, portanto, não foi necessário refinamento das alças. O alinhamento molecular das estruturas resultou em um valor de RMSD = 0,17 Å entre os C α das estruturas, sugerindo que a estrutura construída é muito similar à estrutura molde e, portanto, de boa qualidade.

Em comparação com a estrutura molde, o modelo construído apresentou uma cavidade de ligação um pouco mais fechada devido à conformação dos resíduos Phe105 e Tyr211, o que poderia interferir nos processos de docagem molecular posteriormente desenvolvidos. Portanto, a conformação dos resíduos Phe105 e Tyr211 foi revisada com base na biblioteca de rotâmeros para representar a cavidade de ligação mais apropriada para os derivados benzotiofênicos.

4.4 Construção dos modelos QSAR 3D

Os métodos de QSAR 3D são totalmente dependentes da conformação das moléculas e da distribuição de cargas atômicas do conjunto de dados. Por esse motivo, as etapas de cálculo de cargas atômicas parciais, minimização da conformação estrutural e alinhamento molecular são determinantes da qualidade dos modelos gerados. Diferentes tipos de cargas atômica e alinhamento molecular foram desenvolvidos com o conjunto de dados dos compostos

benzotiofênicos, buscando-se a obtenção de modelos CoMFA e CoMSIA com melhor qualidade preditiva.

Inicialmente, foram construídos diferentes modelos a partir do conjunto de dados da *Pf*NMT com variação do tipo de carga e de alinhamento, sendo baseados em parâmetros padrão dos métodos CoMFA e CoMSIA. Com isso, foi possível determinar os melhores modelos, que posteriormente foram otimizados com a variação dos parâmetros dos métodos de QSAR 3D.

4.4.1 Cálculo de cargas atômicas parciais

Em virtude de as cargas parciais atômicas serem utilizadas para calcular os campos eletrostáticos nos estudos de QSAR 3D, os métodos de cálculo de cargas escolhidos influenciam nos campos eletrostáticos calculados e, conseqüentemente, na qualidade dos modelos gerados.

As cargas empíricas Gasteiger-Hückel e MMFF94, a carga semi-empírica AM1-BCC e as cargas de métodos quânticos CHELPG e Mulliken foram calculadas para cada um dos compostos benzotiofênicos do conjunto de dados. A Figura 20 representa a estrutura molecular do composto **36** após o cálculo das cargas e otimização estrutural, comparando as diferentes conformações e distribuições de carga do composto para os métodos Gasteiger-Hückel, MMFF94, AM1-BCC, CHELPG e Mulliken.

Figura 20 - Representação da estrutura molecular do composto **36** na conformação de menor energia para os diferentes tipos de carga atômica parcial calculados. (A) Carga Gasteiger-Hückel; (B) Carga MMFF94; (C) Carga AM1-BCC; (D) Carga CHELPG; (E) Carga Mulliken. O espectro de cores (azul, branco e vermelho) representam a distribuição de cargas parciais atômicas do composto

Fonte: Elaborada pela autora

4.4.2 Alinhamento molecular

O alinhamento molecular dos compostos é etapa crucial ao desenvolvimento dos modelos QSAR 3D. A distância relativa entre a sonda molecular e os átomos dos compostos depende da orientação das moléculas dentro da grade tridimensional reticulada, o que influencia o cálculo das propriedades estereoquímicas, eletrostáticas, hidrofóbicas e de ligação de hidrogênio.

Neste estudo, três formas de alinhamento molecular foram aplicadas: o alinhamento baseado em MCS, o alinhamento flexível de ligantes e o alinhamento baseado na estrutura do alvo molecular. O alinhamento baseado em MCS foi desenvolvido a partir da sobreposição dos átomos de carbono do anel benzotiofeno - anel benzofurano, no caso dos compostos **1** e **2** – de cada uma das moléculas do conjunto de dados. A Figura 21 mostra o alinhamento dos compostos na conformação minimizada para cada uma das cargas parciais calculadas.

Figura 21 - Representação do alinhamento estrutural baseado na MCS dos compostos na conformação de menor energia para os diferentes tipos de carga atômica parcial calculados. (A) Carga Gasteiger-Hückel; (B) Carga MMFF94; (C) Carga AM1-BCC; (D) Carga CHelpG; (E) Carga Mulliken

Fonte: Elaborada pela autora

O alinhamento flexível de ligantes foi desenvolvido buscando os confôrmeros dos ligantes mais próximos à molécula referência **34** cristalográfica (PDB 4CAF). A Figura 22

mostra o alinhamento dos compostos para cada um dos conjuntos com diferentes cargas parciais calculadas.

No alinhamento baseado na estrutura do alvo molecular, os compostos foram orientados na cavidade de ligação do modelo estrutural da enzima *Pf*NMT construída previamente a partir da modelagem por homologia. Para cada composto do conjunto de dados, a pose (conformação e posição espacial) com melhor pontuação foi selecionada para a construção dos modelos QSAR 3D. A Figura 23 mostra o alinhamento dos compostos, que, nesse caso, é independente da carga atômica calculada.

Figura 22 - Representação do alinhamento flexível de ligantes dos compostos para os diferentes tipos de carga atômica parcial calculados. (A) Carga Gasteiger-Hückel; (B) Carga MMFF94; (C) Carga AM1-BCC; (D) Carga CHELPG; (E) Carga Mulliken

Fonte: Elaborada pela autora

Figura 23 - Representação dos alinhamentos estruturais baseado na estrutura do alvo molecular (*PfNMT*) desenvolvidos pela docagem molecular dos compostos com os programas (A) Surflex-Dock e (B) Glide

Fonte: Elaborada pela autora

4.4.3 Modelos CoMFA e CoMSIA para o conjunto de dados da *PfNMT*

Diferentes metodologias de alinhamento molecular e de cálculo de cargas foram desenvolvidas, buscando a identificação de descritores 3D que auxiliem na melhor interpretação da relação estrutura-atividade dessa série, bem como no desenvolvimento de novos derivados mais potentes e seletivos para a enzima homóloga do parasita. Os estudos de CoMFA e CoMSIA foram realizados para o conjunto de dados da *PfNMT* com os métodos de cálculo de carga Gasteiger-Hückel, MMFF94, AM1-BCC, CHELPG e Mulliken e com os métodos de alinhamento molecular MCS e alinhamento flexível de ligante e docagem no alvo molecular (programas Glide e Surflex-Dock).

Os descritores moleculares foram gerados pelos métodos CoMFA e CoMSIA utilizando os parâmetros padrão. A divisão do conjunto de dados entre conjunto treinamento e teste foi a mesma utilizada na construção dos modelos HQSAR. A fim de encontrar possíveis *outliers*, cada um dos compostos foi excluído do conjunto treinamento e os parâmetros estatísticos dos novos modelos construídos foram analisados. O composto **10** foi determinado *outlier* nos modelos gerados e, por isso, foi excluído da construção de todos os modelos. A contribuição estérica dos modelos CoMSIA foi desconsiderada, dado que tinha valor fracionário menor que 0,1 para todos os modelos gerados.

Os valores de pK_i foram utilizados como variável dependente e correlacionados com os descritores moleculares dos métodos CoMFA e CoMSIA (variável independente) utilizando o método de SAMPLS. Os modelos tiveram a capacidade preditiva validada pelo método de

validação cruzada LOO. A Tabela 12 e a Tabela 13 apresentam os resultados estatísticos dos modelos obtidos.

Tabela 12 - Resultados estatísticos dos modelos QSAR 3D CoMFA iniciais para os inibidores benzotiofênicos contra a enzima *Pf*NMT

Modelo	Carga Atômica	Alinhamento Molecular	N	q^2	SEP	r^2	SEE
1		MCS	2	0,13	0,94	0,68	0,57
2	Gasteiger-	Flexível	4	0,51	0,73	0,97	0,18
3	Hückel	Glide	3	0,37	0,84	0,79	0,48
4		Surflex-Dock	6	0,27	0,93	0,99	0,13
5		MCS	6	0,53	0,75	1,00	0,06
6	MMFF94	Flexível	4	0,61	0,65	0,98	0,16
7		Glide	5	0,16	0,97	0,98	0,14
8		Surflex-Dock	5	0,31	0,89	0,97	0,18
9		MCS	3	0,54	0,69	0,97	0,19
10	AM1-BCC	Flexível	3	0,56	0,68	0,96	0,20
11		Glide	2	0,04	0,98	0,82	0,42
12		Surflex-Dock	4	0,38	0,82	0,97	0,19
13		MCS	1	0,35	0,79	0,70	0,54
14	CHELPG	Flexível	4	0,64	0,62	0,97	0,18
15		Glide	6	0,38	0,85	0,98	0,15
16		Surflex-Dock	4	0,34	0,85	0,98	0,16
17		MCS	1	0,43	0,74	0,74	0,50
18	MULLIKEN	Flexível	5	0,62	0,66	0,99	0,12
19		Glide	6	0,30	0,91	0,98	0,15
20		Surflex-Dock	4	0,31	0,87	0,96	0,20

N, número ótimo de componentes principais; q^2 , coeficiente de correlação da validação cruzada; SEP, erro padrão da predição; r^2 , coeficiente de correlação; SEE, erro padrão da correlação

Fonte: Elaborada pela autora

Tabela 13 - Resultados estatísticos dos modelos QSAR 3D CoMSIA iniciais para os inibidores benzotiofênicos contra a enzima *Pf*NTMT

Modelo	Carga Atômica	Alinhamento Molecular	N	q^2	SEP	r^2	SEE
21		MCS	4	0,64	0,63	0,97	0,19
22	Gasteiger-	Flexível	2	0,44	0,75	0,84	0,40
23	Hückel	Glide	6	0,33	0,93	0,94	0,23
24		Surflex-Dock	1	0,43	0,74	0,71	0,53
25		MCS	3	0,69	0,57	0,93	0,28
26	MMFF94	Flexível	4	0,60	0,66	0,93	0,27
27		Glide	5	0,19	0,80	0,87	0,37
28		Surflex-Dock	1	0,43	0,74	0,71	0,53
29		MCS	3	0,65	0,61	0,92	0,28
30	AM1-BCC	Flexível	5	0,55	0,71	0,99	0,13
31		Glide	4	0,35	0,68	0,97	0,23
32		Surflex-Dock	1	0,42	0,75	0,72	0,52
33		MCS	3	0,72	0,55	0,96	0,22
34	CHELPG	Flexível	3	0,73	0,54	0,95	0,23
35		Glide	5	0,34	0,86	0,98	0,17
36		Surflex-Dock	1	0,41	0,76	0,70	0,54
37		MCS	3	0,71	0,55	0,96	0,22
38	MULLIKEN	Flexível	5	0,70	0,58	0,98	0,16
39		Glide	5	0,32	0,92	0,97	0,18
40		Surflex-Dock	2	0,40	0,78	0,84	0,40

N, número ótimo de componentes principais; q^2 , coeficiente de correlação da validação cruzada; SEP, erro padrão da predição; r^2 , coeficiente de correlação; SEE, erro padrão da correlação

Fonte: Elaborada pela autora

Como pode ser observado nas tabelas, no geral, os modelos CoMSIA apresentaram melhores parâmetros estatísticos (r^2 e q^2 mais altos) em relação ao CoMFA, o que sugere que os cálculos de potenciais desenvolvidos nos modelos CoMSIA conseguiram representar melhor as moléculas do conjunto de dados para a construção da correlação com a atividade biológica. A partir da análise desses dados, é possível notar que os modelos que utilizaram alinhamento baseado na estrutura do ligante apresentaram melhores resultados em comparação aos modelos

baseados na estrutura do alvo molecular. Os modelos construídos a partir das moléculas alinhadas pelo método *Flexible Ligand Alignment* do programa Schrödinger, que utiliza a busca de confômeros e teve como estrutura referência uma molécula cristalográfica (**34**, PDB 4CAF), tiveram os melhores resultados se comparados aos modelos que utilizaram os outros métodos de alinhamento estrutural. Por apresentar os melhores resultados, considera-se que esse método de alinhamento deve melhor representar a estrutura molecular dos compostos na conformação biologicamente ativa.

Considerando os modelos gerados pelo alinhamento flexível, as cargas parciais atômicas calculadas pelos métodos quânticos (CHELPG e Mulliken) resultaram em modelos estatisticamente mais preditivos com valores ótimos de parâmetros estatísticos ($r^2 > 0,9$, $q^2 > 0,6$). Seguido dos modelos com cargas quânticas calculadas, o modelo com carga empírica MMFF94 apresentou também ótimos resultados, dado que requer menor custo de máquina para realizar os cálculos. Esses resultados sugerem que as cargas calculadas pelos métodos quânticos foram mais eficientes em representar a distribuição eletrônica dos ligantes para a construção dos modelos de QSAR 3D.

Entre esses modelos gerados inicialmente, os modelos 14 (CoMFA, carga quântica CHELPG, alinhamento flexível) e 34 (CoMSIA, carga quântica CHELPG, alinhamento flexível) foram selecionados como melhores modelos para as próximas etapas, de acordo com os parâmetros estatísticos, priorizando maiores valores de r^2 e q^2 e menores valores de erro associado (SEE e SEP). O modelo CoMFA foi otimizado utilizando o procedimento de focagem de região, que atribui seletivamente novos pesos aos valores de potencial de cada ponto de intersecção do *grid*. Dois processos foram utilizados para derivar os novos pesos: o *Discriminant Power*, que atua de acordo com a variação da fração dos componentes em cada ponto de intersecção; e o *Standard Deviation Coefficient* ($Stdev * Coefficients$), que atua de acordo com o produto da variação em cada ponto e o coeficiente de regressão. Diferentes pesos foram analisados para cada um dos processos e o espaçamento entre pontos de intersecção do *grid* também foi variado (2 e 1 Å) buscando a otimização dos modelos. A Tabela 14 apresenta os resultados de parâmetros estatísticos obtidos para o refinamento do melhor modelo CoMFA.

Tabela 14 - Análise dos modelos CoMFA com base nos parâmetros estatísticos para diferentes definições de espaçamento do *grid* e região de focagem para o conjunto de dados da *PfNMT*

Modelo	Peso	Espaçamento de <i>grid</i> (Å)	N	q^2	SEP	r^2	SEE
1	Std 0,3	2	4	0,76	0,51	0,97	0,17
2	Std 0,5		4	0,77	0,50	0,97	0,19
3	Std 0,8		4	0,75	0,52	0,96	0,22
4	Std 1,2		3	0,69	0,57	0,92	0,28
5	Std 1,5		2	0,64	0,60	0,83	0,41
6	Std 0,3	1	4	0,70	0,57	0,98	0,16
7	Std 0,5		4	0,71	0,56	0,97	0,17
8	Std 0,8		4	0,71	0,56	0,97	0,19
9	Std 1,2		4	0,68	0,59	0,96	0,20
10	Std 1,5		4	0,65	0,62	0,96	0,22
11	Disc 0,3	2	4	0,72	0,55	0,97	0,19
12	Disc 0,5		5	0,72	0,57	0,97	0,18
13	Disc 0,8		5	0,71	0,57	0,96	0,21
14	Disc 1,2		5	0,70	0,58	0,95	0,24
15	Disc 1,5		4	0,69	0,58	0,92	0,30
16	Disc 0,3	1	4	0,66	0,61	0,97	0,18
17	Disc 0,5		4	0,66	0,61	0,97	0,18
18	Disc 0,8		4	0,61	0,65	0,95	0,23
19	Disc 1,2		5	0,57	0,69	0,97	0,20
20	Disc 1,5		4	0,56	0,69	0,93	0,28

N, número ótimo de componentes principais; q^2 , coeficiente de correlação da validação cruzada; SEP, erro padrão da predição; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; Std, *Stdev*Coefficients*; Disc, *Discriminant Power*

Fonte: Elaborada pela autora

A tabela mostra que o procedimento de focagem de região conseguiu refinar o modelo CoMFA selecionado, dado que os parâmetros coeficientes de correlação obtidos foram mais altos. O modelo 2 construído pelo procedimento de focagem de região *Standard Deviation Coefficient* com peso 0,5 e mantendo o espaçamento de *grid* padrão de 2 Å foi selecionado como modelo de maior consistência interna.

Com isso, o melhor modelo CoMFA – construído com as moléculas alinhadas pelo método *Flexible Ligand Alignment*, com cargas parciais atômicas CHELPG, focagem de região *Standard Deviation Coefficient* com peso 0,5 e espaçamento de *grid* 2 Å – e o melhor CoMSIA – construído com as moléculas alinhadas pelo método *Flexible Ligand Alignment* e com cargas parciais atômicas CHELPG – apresentaram os coeficientes de correlação $q^2 = 0,77/ r^2 = 0,97$ e $q^2 = 0,73/ r^2 = 0,95$, respectivamente. As frações de contribuição dos campos potenciais para o modelo CoMFA foram de 0,45 de contribuição estérica e 0,55 de contribuição eletrostática. No caso do modelo CoMSIA, as contribuições foram 0,38 de contribuição eletrostática, 0,27 de contribuição de aceptores de ligação de hidrogênio, 0,18 de contribuição de doador de ligação de hidrogênio e 0,17 de contribuição hidrofóbica.

Esses modelos CoMFA e CoMSIA foram selecionados para a etapa de validação externa, na qual cada um calculou a predição de pK_i das moléculas previamente selecionadas para o conjunto teste. A Tabela 15 e a Tabela 16 apresentam os resultados da predição do conjunto teste para os modelos CoMFA e CoMSIA, respectivamente. A Figura 24 mostra o gráfico dos valores de pK_i experimentais *versus* preditos para ambos os métodos.

Tabela 15 - Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo CoMFA da *PfNMT*

Composto	pK_i		
	Experimental	Predito	Resíduo*
5	6,08	6,21	0,13
12	5,95	6,08	0,13
14	5,00	5,56	0,56
16	7,15	7,23	0,08
23	7,82	7,39	-0,43
26	7,48	7,14	-0,34
35	6,95	7,35	0,40
37	6,65	6,93	0,28

*Resíduo = pK_i experimental – pK_i predito

Fonte: Elaborada pela autora

Tabela 16 - Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo CoMSIA da PfNMT

Composto	pK_i		
	Experimental	Predito	Resíduo*
5	6,08	6,06	-0,02
12	5,95	6,25	0,30
14	5,00	5,09	0,09
16	7,15	7,31	0,16
23	7,82	7,05	-0,77
26	7,48	7,04	-0,44
35	6,95	7,25	0,30
37	6,65	7,07	0,42

*Resíduo = pK_i experimental – pK_i predito

Fonte: Elaborada pela autora

Os modelos CoMFA e CoMSIA apresentaram ótimos resultados de validação externa. A distribuição do gráfico de pK_i experimental *versus* predito mostra a boa correlação dos dados, tanto internos à construção dos modelos (pontos pretos) quanto externos (pontos vermelhos). Além disso, 85% dos valores de predição de atividade biológica apresentaram resíduo em relação aos valores experimentais menores que 0,5 em escala logarítmica. A partir dos valores preditos de pK_i , foram calculados os coeficientes r^2_{pred} . Os valores obtidos foram muito satisfatórios, o modelo CoMFA apresentou o valor de $r^2_{pred} = 0,86$ e o modelo CoMSIA, $r^2_{pred} = 0,82$, o que indica que esses modelos têm uma ótima capacidade preditiva para novas moléculas estruturalmente relacionadas ao conjunto de dados dos benzotiofenos.

Além das predições, esses modelos de QSAR 3D geram mapas de contorno, que sintetizam visualmente quais são as modificações estruturais sugeridas para a otimização da atividade biológica. Os mapas de contorno do CoMFA mostram as características estéreas – regiões em verde devem ter substituintes volumosos e regiões em amarelo não devem ter substituintes volumosos – e eletrostáticas – regiões em vermelho são favoráveis para substituintes eletronegativos e regiões em azul são favoráveis para substituintes eletropositivos.

Os mapas de contorno do CoMSIA apresentam as características hidrofóbicas – regiões em amarelo, favoráveis para substituintes hidrofóbicos e regiões em branco, desfavoráveis para substituintes hidrofóbicos -, eletrostáticas - regiões em vermelho, favoráveis para substituintes eletronegativos e regiões em azul, favoráveis para substituintes eletropositivos – e de doadores e aceptores de ligação de hidrogênio – favoráveis para doadores,

em azul ciano; desfavoráveis para doadores, em roxo; favoráveis para aceptores, em magenta; desfavoráveis para aceptores, em vermelho. A Figura 25 e a Figura 26 apresentam os mapas de contorno gerados pelos modelos CoMFA e CoMSIA, respectivamente.

A

B

Figura 24 - Representação gráfica dos valores experimentais *versus* preditos de pK_i para os melhores modelos (**A**) CoMFA e (**B**) CoMSIA do conjunto de dados da *PfNMT*, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos

Fonte: Elaborada pela autora

Figura 25 - Mapas de contorno gerados para o melhor modelo CoMFA a partir do conjunto de dados da *Pf*NMT, sendo o composto mais potente (**36**) a molécula de referência. (A) Mapa de contorno estérico, representado em verde e amarelo (níveis de contorno 0,005 e -0,001 kcal/mol, respectivamente). (B) Mapa de contorno eletrostático, representado em azul e vermelho (níveis de contorno 0,10 e -0,05 kcal/mol, respectivamente)

Fonte: Elaborada pela autora

Figura 26 - Mapas de contorno gerados para o melhor modelo CoMSIA a partir do conjunto de dados da *Pf*NMT sendo o composto mais potente (**36**) a molécula de referência. (A) Mapa de contorno hidrofóbico representado em amarelo e branco (níveis de contorno 0,005 e -0,010 kcal/mol, respectivamente). (B) Mapa de contorno eletrostáticos representado em azul e vermelho (níveis de contorno 0,003 e -0,003 kcal/mol, respectivamente). (C) Mapa de contorno de doadores, representado em ciano e roxo (níveis de contorno 0,01 e -0,05 kcal/mol, respectivamente) e aceptores de ligação de hidrogênio, representado em magenta e vermelho (níveis de contorno 0,05 e -0,03 kcal/mol, respectivamente)

Fonte: Elaborada pela autora

O mapa de contorno estéreo gerado pelo melhor modelo CoMFA indica que substituintes volumosos (cor verde) nas posições 3 e 5 do anel pirazol do composto **36** teriam influência positiva na atividade biológica. Já o mapa eletrostático gerado por esse modelo, sugere substituintes eletronegativos (cor vermelha) na posição 3 do anel pirazol.

O mapa de contorno hidrofóbico gerado pelo melhor modelo CoMSIA concorda com o mapa estéreo do CoMFA sobre as substituições nas posições 3 e 5 do anel pirazol, portanto, o modelo sugere que esses substituintes deveriam ter característica hidrofóbica. O mapa eletrostático do CoMSIA apresentou contornos em regiões muito próximas aos contornos do mapa eletrostático do modelo CoMFA. Esse mapa sugere substituições de caráter eletronegativo (cor vermelha) na região da posição 3 do anel pirazol. O mapa de ligações de hidrogênio indica que aceptores (cor magenta) na posição 5 do anel pirazol poderiam aumentar a afinidade do composto **36**.

4.4.4 Modelos CoMFA e CoMSIA para o conjunto de dados da *Hs*NMT

A partir dos resultados obtidos para o conjunto de dados da enzima *Pf*NMT, foram selecionados os melhores métodos de alinhamento molecular e de cálculo de cargas parciais atômicas para serem aplicados ao conjunto de dados da enzima *Hs*NMT. Portanto, as moléculas do conjunto de dados da enzima *Hs*NMT para construção dos modelos CoMFA e CoMSIA foram alinhadas pelo método *Flexible Ligand Alignment* e tiveram as cargas CHELPG calculadas.

A divisão das moléculas entre conjunto treinamento e teste foi a mesma utilizada na construção do modelo HQSAR para a *Hs*NMT. Inicialmente, os modelos foram gerados com parâmetros padrão para determinar os descritores moleculares. Por exclusão de cada um dos compostos do conjunto de dados, duas moléculas foram identificadas como *outliers* e excluídas do conjunto treinamento (compostos **6** e **10**). Por apresentar valor fracionário menor que 0,1, a contribuição estérea do modelo CoMSIA foi desconsiderada na construção desse modelo. A Tabela 17 sumariza os resultados obtidos.

Os valores dos parâmetros estatísticos de correlação para ambos os modelos foram satisfatórios ($q^2 > 0,5$ e $r^2 > 0,6$). O modelo CoMFA ainda foi otimizado utilizando o procedimento de focagem de região. A Tabela 18 apresenta os resultados de refinamento para o modelo CoMFA para os diferentes processos de refinamento (*Discriminant Power* e *Standard Deviation Coefficient*) e espaçamento de *grid*.

Tabela 17 - Resultados estatísticos dos modelos QSAR 3D, CoMFA e CoMSIA, para os inibidores benzotiofênicos contra a enzima *HsNMT*

Modelo	N	q^2	SEP	r^2	SEE
CoMFA	4	0,60	0,63	0,98	0,15
CoMSIA	2	0,61	0,65	0,95	0,22

N, número ótimo de componentes principais; q^2 , coeficiente de correlação da validação cruzada; SEP, erro padrão da predição; r^2 , coeficiente de correlação; SEE, erro padrão da correlação

Fonte: Elaborada pela autora

Os resultados mostram que o procedimento de focagem de região foi capaz de otimizar o modelo CoMFA referente ao conjunto de dados da enzima *HsNMT*. O modelo 2, construído pelo procedimento de focagem de região *Standard Deviation Coefficient* com peso 0,5 e mantendo o espaçamento de *grid* padrão de 2 Å, foi selecionado como modelo de maior consistência interna para posterior etapa de validação.

Os melhores modelos de QSAR 3D referentes ao conjunto de dados da *HsNMT* foram: o modelo CoMFA construído com as moléculas alinhadas pelo método *Flexible Ligand Alignment*, com cargas parciais atômicas CHELPG, focagem de região *Standard Deviation Coefficient* com peso 0,5 e espaçamento de *grid* 2 Å, e o modelo CoMSIA construído com as moléculas alinhadas pelo método *Flexible Ligand Alignment* e com cargas parciais atômicas CHELPG. Esses modelos apresentaram os coeficientes de correlação $q^2 = 0,81$ / $r^2 = 0,98$ e $q^2 = 0,61$ / $r^2 = 0,95$, respectivamente. As frações de contribuição dos campos potenciais para o modelo CoMFA foram de 0,48 de contribuição estérica e 0,52 de contribuição eletrostática. No caso do modelo CoMSIA, as contribuições foram 0,41 de contribuição eletrostática, 0,27 de contribuição de aceptores de ligação de hidrogênio, 0,13 de contribuição de doador de ligação de hidrogênio e 0,19 de contribuição hidrofóbica.

Esses modelos foram selecionados para a etapa de validação externa, na qual calcularam a predição de pK_i das moléculas previamente selecionadas para o conjunto teste. A Tabela 19 e a Tabela 20 apresentam os resultados da predição do conjunto teste para os modelos CoMFA e CoMSIA, respectivamente. A Figura 27 mostra a plotagem dos valores de pK_i experimentais *versus* preditos para ambos os métodos.

Tabela 18 - Análise dos modelos CoMFA com base nos parâmetros estatísticos para diferentes definições de espaçamento do *grid* e região de focagem para o conjunto de dados da *HsNMT*

Modelo	Peso	Espaçamento de <i>grid</i> (Å)	N	q^2	SEP	r^2	SEE
1	Std 0,3	2	4	0,78	0,46	0,98	0,15
2	Std 0,5		4	0,81	0,44	0,98	0,16
3	Std 0,8		4	0,81	0,43	0,97	0,19
4	Std 1,2		4	0,80	0,44	0,96	0,21
5	Std 1,5		4	0,78	0,46	0,95	0,22
6	Std 0,3	1	4	0,62	0,61	0,98	0,15
7	Std 0,5		4	0,66	0,58	0,98	0,15
8	Std 0,8		4	0,68	0,56	0,98	0,15
9	Std 1,2		4	0,70	0,55	0,97	0,18
10	Std 1,5		4	0,70	0,54	0,96	0,20
11	Disc 0,3	2	4	0,74	0,51	0,97	0,18
12	Disc 0,5		4	0,79	0,46	0,96	0,19
13	Disc 0,8		4	0,81	0,44	0,95	0,23
14	Disc 1,2		4	0,81	0,43	0,93	0,26
15	Disc 1,5		4	0,80	0,44	0,92	0,28
16	Disc 0,3	1	4	0,57	0,65	0,97	0,17
17	Disc 0,5		4	0,61	0,62	0,97	0,18
18	Disc 0,8		4	0,62	0,61	0,96	0,20
19	Disc 1,2		4	0,65	0,59	0,94	0,25
20	Disc 1,5		4	0,66	0,58	0,93	0,27

N, número ótimo de componentes principais; q^2 , coeficiente de correlação da validação cruzada; SEP, erro padrão da predição; r^2 , coeficiente de correlação; SEE, erro padrão da correlação; Std, *Stdev*Coefficients*; Disc, *Discriminant Power*

Fonte: Elaborada pela autora

Os valores de pK_i preditos pelos modelos mostraram-se razoavelmente próximos aos valores experimentais. Os gráficos de pK_i experimental *versus* predito apresentam melhor correlação no intervalo de pK_i de 5,5 a 7,5. Os ligantes com predições menos precisas foram os compostos **3** e **23**, que estão nos extremos da distribuição de pK_i . O valor do coeficiente de predição calculado para os modelos foi de $r^2_{pred} = 0,60$ para o modelo CoMFA e $r^2_{pred} = 0,56$ para o modelo CoMSIA.

Tabela 19 - Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo CoMFA da *HsNMT*

Composto	pK_i		
	Experimental	Predito	Resíduo*
3	4,00	5,00	1,00
12	5,73	5,55	-0,18
19	6,06	6,52	0,46
23	7,30	6,45	-0,85
26	6,37	5,91	-0,46
29	6,01	5,63	-0,38
34	5,81	6,42	0,61

*Resíduo = pK_i experimental – pK_i predito

Fonte: Elaborada pela autora

Tabela 20 - Valores experimentais e preditos de pK_i dos compostos do conjunto teste obtidos pelo melhor modelo CoMSIA da *HsNMT*

Composto	pK_i		
	Experimental	Predito	Resíduo*
3	4,00	5,27	1,27
12	5,73	5,58	-0,16
19	6,06	6,43	0,37
23	7,30	6,50	-0,80
26	6,37	6,13	-0,24
29	6,01	5,86	-0,15
34	5,81	6,51	0,70

*Resíduo = pK_i experimental – pK_i predito

Fonte: Elaborada pela autora

Figura 27 - Representação gráfica dos valores experimentais *versus* preditos de pK_i para os melhores modelos (A) CoMFA e (B) CoMSIA do conjunto de dados da *HsNMT*, com o conjunto treinamento representado por pontos pretos e o conjunto teste, por pontos vermelhos

Fonte: Elaborada pela autora

Os mapas de contorno gerados pelos modelos CoMFA e CoMSIA da enzima *HsNMT* foram comparados com os mapas de contorno construídos a partir do conjunto de dados do parasita. Essa análise sugere modificações na estrutura dos derivados benzotiofênicos que devem favorecer a atividade biológica contra a enzima do parasita (*PfNMT*) e desfavorecer a inibição da homóloga humana (*HsNMT*). A Figura 28 e a Figura 29 apresentam os mapas de contorno referentes aos modelos CoMFA e CoMSIA, respectivamente, comparando os mapas da *PfNMT* e da *HsNMT*.

A

B

Figura 28 - Mapas de contorno gerados para os melhores modelos CoMFA a partir do conjunto de dados da *Pj*NMT (à esquerda) e *Hs*NMT (à direita), sendo o composto mais potente (**36**) a molécula de referência. (A) Mapa de contorno estérico (B) Mapa de contorno eletrostático

Fonte: Elaborada pela autora

Figura 29 - Mapas de contorno gerados para os melhores modelos CoMSIA a partir do conjunto de dados da *Pf*NMT (à esquerda) e *Hs*NMT (à direita), sendo o composto mais potente (**36**) a molécula de referência. (A) Mapa de contorno hidrofóbico. (B) Mapa de contorno eletrostáticos. (C) Mapa de contorno de doadores e aceptores de ligação de hidrogênio

Fonte: Elaborada pela autora

Como pode ser analisado pela comparação dos mapas de contorno correspondentes às enzimas *Pf*NMT e *Hs*NMT, a maioria das regiões coloridas são comuns a ambos os mapas. No entanto, os mapas estéreo do CoMFA mostram contornos verdes maiores para a *Pf*NMT e o mapa hidrofóbico do CoMSIA para a *Pf*NMT apresenta regiões favoráveis para substituintes hidrofóbicos (contorno amarelo) que não existem no mapa correspondente da *Hs*NMT. Além disso, o mapa de contorno do CoMSIA de ligações de hidrogênio da *Hs*NMT não indicou uma das regiões favoráveis a aceptores que está presente no mapa da *Pf*NMT. Portanto, as substituições sugeridas pela comparação dos mapas de contorno para que haja maior seletividade para a enzima do parasita são: substituintes hidrofóbicos nas posições 3 e 5 do anel pirazol e substituintes aceptores de ligação de hidrogênio na posição 5 do anel pirazol.

4.5 Planejamento de novos candidatos a fármacos antimaláricos

Seguido da análise dos mapas de contribuição dos modelos HQSAR e mapas de contorno dos modelos CoMFA e CoMSIA, foi possível notar concordância e complementariedade entre eles. Comparando-se os mapas dos modelos criados a partir do conjunto de dados da NMT parasita e da NMT humana, informações importantes foram obtidas para o desenvolvimento de compostos com maior afinidade e seletividade para a enzima do parasita. No geral, grupos volumosos e hidrofóbicos (terc-butil, ciclopropil, fenil) - nas posições 3 e 5 do anel pirazol devem aumentar a atividade biológica do composto **36**, assim como substituintes eletronegativos (F, Br, Cl, OH) na posição 3 e aceptores de ligação de hidrogênio (carbonila, O, N) na posição 5 do anel pirazol. Em busca de maior seletividade, as modificações estruturais indicadas pelos mapas são: grupos hidrofóbicos nas posições 3 e 5 do anel pirazol, grupos aceptores de ligação de hidrogênio na posição 5 do anel pirazol e mudança nos átomos das posições 4 e 5 do benzeno do anel benzotiofeno, dado que os mapas de contribuição apontam a influência positiva desses átomos apenas contra a enzima humana. A Figura 30 sumariza essas modificações.

Figura 30 – Representação das modificações estruturais sugeridas para o composto **36** pelos modelos HQSAR, CoMFA e CoMSIA
 Fonte: Elaborada pela autora

CONCLUSÕES E PERSPECTIVAS

5 CONCLUSÕES E PERSPECTIVAS

Nesta dissertação de mestrado foram desenvolvidos estudos de modelagem molecular que incluíram a modelagem por homologia da enzima alvo e o estudo de relação quantitativa entre estrutura e atividade biológica buscando-se a identificação das bases moleculares envolvidas no processo de reconhecimento molecular entre ligantes (derivados benzotiofênicos) e alvo biológico (*Pf*NMT) para o desenvolvimento de novos candidatos a fármacos antimaláricos. Métodos de QSAR 2D (HQSAR) e QSAR 3D (CoMFA e CoMSIA) foram aplicados sobre os conjuntos de dados referentes a atividade biológica contra a enzima do parasita (*P. falciparum*) e a enzima humana. Esses estudos permitiram a criação de hipóteses sobre modificações estruturais para maior afinidade e seletividade pela enzima *Pf*NMT. Nos métodos CoMFA e CoMSIA, foram utilizadas diversas estratégias de alinhamento molecular e cálculo de cargas parciais atômicas, possibilitando a identificação e seleção dos parâmetros mais apropriados para a modelagem das estruturas e atividade biológica das moléculas dos conjuntos de dados. Dentre os resultados obtidos, destacam-se:

- Os modelos HQSAR criados para os conjuntos de dados da *Pf*NMT e da *Hs*NMT tiveram resultados estatísticos de qualidade, indicando a boa capacidade preditiva para novos compostos;
- Os mapas de contribuição do HQSAR forneceram informações relevantes sobre a relação estrutura-atividade dos compostos benzotiofênicos;
- O modelo estrutural da *Pf*NMT, obtido pelo método de modelagem por homologia, foi construído com sucesso, o que permitiu a docagem dos compostos benzotiofênicos nesse alvo molecular;
- Os modelos QSAR 3D (CoMFA e CoMSIA) com diferentes métodos de cálculo de carga e alinhamento molecular foram construídos para o conjunto de dados da *Pf*NMT;
- Os modelos CoMFA e CoMSIA para os conjuntos de dados da *Pf*NMT e da *Hs*NMT apresentaram resultados com boa consistência interna e externa, indicando sua qualidade em prever a atividade biológica de novos ligantes;
- Os mapas de contorno dos modelos CoMFA e CoMSIA indicaram regiões para modificações estruturais visando maior afinidade e seletividade para novos compostos;
- As informações estruturais obtidas neste estudo são úteis para guiar a síntese de novos derivados com maior afinidade e seletividade como candidatos a novos fármacos antimaláricos.

REFERÊNCIAS

- 1 WALKER, N. F.; NADJM B.; WHITTY C. J. M. Malaria. **Medicine** (Baltimore), v. 38 n. 1 p. 41-46; 2010. doi: 10.1016/j.mpmed.2009.09.017.
- 2 MENARD, R. et al. Looking under the skin: the first steps in malarial infection and immunity. **Nature Reviews Microbiology**, v. 11, p. 701-712, 2013. doi: 10.1038/nrmicro3111.
- 3 FLANNERY, E. L.; CHATTERJEE, A. K.; WINZELER, E. A. Antimalarial drug discovery - approaches and progress towards new medicines. **Nature Reviews Microbiology**, v. 11, n. 12, p. 849-862, 2013. doi: 10.1038/nrmicro3138.
- 4 MILLER, L. H.; ACKERMAN, H.C.; SU, X.; WELLEMS. T. E. Malaria biology and disease pathogenesis: insights for new treatments. **Nature Medicine**, v. 19, n. 2 p. 156-167, 2013. doi: 10.1038/nm.3073.
- 5 GAZZINELLI, R. T. et al. Innate sensing of malaria parasites. **Nature Reviews Immunology**, v. 14, p. 744-757, 2014. doi: 10.1038/nri3742.
- 6 GREENWOOD, B. M. et al. Malaria. **Lancet**, v. 365, n. 9469, p. 1487-1498, 2005. doi: 10.1016/S0140-6736(05)66420-3.
- 7 STORM, J.; CRAIG, A. G. Pathogenesis of cerebral malaria - inflammation and cytoadherence. **Frontiers in Cellular and Infection Microbiology**, v. 4, n. 100, p. 1-8, 2014. doi: 10.3389/fcimb.2014.00100.
- 8 BREMAN, J. G. The ears of the hippopotamus: Manifestations; determinants; and estimates of the malaria burden. **American Journal of Tropical Medicine and Hygiene**, v. 64, n. 1-2 supl., p. 1-11, 2000. doi: 10.4269/ajtmh.2001.64.1.
- 9 BHUTTA, Z. A. et al. Global burden; distribution; and interventions for infectious diseases of poverty. **Infectious Diseases of Poverty**, v. 3, n. 21, p. 1-7, 2014. doi: 10.1186/2049-9957-3-21.
- 10 SACHS, J.; MALANEY, P. The economic and social burden of malaria. **Nature**, v. 415, n.6872, p. 680-685, 2002. doi: 10.1038/415680a.
- 11 WORLD HEALTH ORGANIZATION. **World Malaria Report 2016**, Geneva, 2016. ISBN: 978-92-4-151171-1.
- 12 NKUMAMA, I. N.; O'MEARA, W. P.; OSIER, F. H. A. Changes in malaria epidemiology in Africa and new challenges for elimination. **Trends in Parasitology**, v. 33, n. 2, p. 128-140, 2016. doi: 10.1016/j.pt.2016.11.006.
- 13 ASHLEY, E. et al. Malaria. **Travel Medicine and Infectious Disease**, v. 4, p. 159-173, 2006. doi: 10.1016/j.tmaid.2005.06.009.
- 14 WORLD HEALTH ORGANIZATION. **Guidelines for the treatment of malaria**. 3rd ed. Geneva, 2015. ISBN: 978-92-4-154912-7.
- 15 DONDORP, A. M. et al. Artemisinin resistance in Plasmodium falciparum malaria. **New England Journal of Medicine**, v. 361, n. 5, p. 455-467, 2009. doi: 10.1056/NEJMoa0808859.
- 16 ROBERT A. et al. From classical antimalarial drugs to new compounds based on the mechanism of action of artemisinin. **Pure and Applied Chemistry**, v. 73, n. 7, p. 1173-1188, 2001. doi: 10.1351/pac200173071173.
- 17 MÜLLER, I. B.; HYDE, J. E. Antimalarial drugs: modes of action and mechanisms of parasite resistance. **Future Microbiology**, v. 5, n. 12, p. 1857-1873, 2010. doi: 10.2217/fmb.10.136.
- 18 MBENGUE, A. et al. A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. **Nature**, v. 520, p. 683-687, 2015. doi: 10.1038/nature14412.
- 19 GEPPERT, T.; KOEPPEN, H. Biological networks and drug discovery-where do we stand? **Drug Development Research**, v. 75, n. 5, p. 271-282, 2014. doi: 10.1002/ddr.21207.

- 20 GUIDO, R. V. C.; OLIVA, G.; ANDRICOPULO, A. D. Modern drug discovery technologies: opportunities and challenges in lead discovery. **Combinatorial Chemistry & High Throughput Screening**, v. 14, n. 10, p. 830-839, 2011. doi: 10.2174/138620711797537067.
- 21 ÖSTER, L. et al. Successful generation of structural information for fragment-based drug discovery. **Drug Discovery Today**, v. 0, n. 0, p. 1-8, 2015. doi: 10.1016/j.drudis.2015.04.005.
- 22 WIGGLESWORTH, M. J. et al. Increasing the delivery of next generation therapeutics from high throughput screening libraries. **Current Opinion in Chemistry Biology**, v. 26, p. 104-110, 2015. doi: 10.1016/j.cbpa.2015.04.006.
- 23 RICHTER, L.; ECKER, G. F. Medicinal chemistry in the era of big data. **Drug Discovery Today Technologies**, v. 14, p. 37-41, 2015. doi: 10.1016/j.ddtec.2015.06.001.
- 24 WERNÉRUS, H.; STÅHL, S. Biotechnological applications for surface-engineered bacteria. **Biotechnology Applied Biochemistry**, v. 40, p. 209-228, 2004. doi: 10.1042/BA20040014.
- 25 LOMBARDINO, J. G.; LOWE, J. A. The role of the medicinal chemist in drug discovery - then and now. **Nature Reviews Drug Discovery**, v. 3, n. 10, p. 853-862, 2004. doi: 10.1038/nrd1523.
- 26 KATSUNO, K. et al. Hit and lead criteria in drug discovery for infectious diseases of the developing world. **Nature Reviews Drug Discovery**, v. 14, n. 11, p. 751-758, 2015. doi: 10.1038/nrd4683.
- 27 HUGHES, J. P. et al. Principles of early drug discovery. **British Journal of Pharmacology**, v. 162, n. 6, p. 1239-1249. doi: 10.1111/j.1476-5381.2010.01127.x.
- 28 MAVROMOUSTAKOS, T. et al. Strategies in the rational drug design. **Current Medicinal Chemistry**, v. 18, n. 17, p. 2517-2530, 2011. doi: 10.2174/092986711795933731.
- 29 FERREIRA, R. S. et al. Screening strategies for novel inhibitors of parasitic diseases. **Expert Opinion on Drug Discovery**, v. 6, n. 5, p. 481-489, 2011. doi: 10.1517/17460441.2011.563297.
- 30 GUIDO, R. V. C.; OLIVA, G.; ANDRICOPULO, A.D. Structure- and ligand-based drug design approaches for neglected tropical diseases. **Pure and Applied Chemistry**, v.84, n. 9, p. 1857-1866, 2012. doi: 10.1351/PAC-CON-11-11-07.
- 31 HUNG, C.; CHEN, C. Computational approaches for drug discovery. **Drug Development Research**, v. 75, n. 6, p. 412-418, 2014. doi: 10.1002/ddr.21222.
- 32 HANSCH, C.; FUJITA, T. ρ - σ - π Analysis. A method for the correlation of biological activity and chemical structure. **Journal of American Chemistry Society**, v. 86, n. 8, p. 1616-1626, 1964. doi: 10.1021/ja01062a035.
- 33 MELO-FILHO, C. C.; BRAGA, R. C.; ANDRADE, C. H. 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. **Current Computer-Aided Drug Design**, v. 10, p. 148-159, 2014. doi: 10.2174/1573409910666140410111043.
- 34 DIVAKAR, S.; HARIHARAN, S. 3D-QSAR studies on Plasmodium falciparum proteins: a mini-review. **Combinatorial Chemistry & High Throughput Screening**, v. 18, p. 188-198, 2015. doi: 10.2174/1386207318666141229124747.
- 35 DUDEK, A. Z.; ARODZ, T.; GÁLVEZ, J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. **Combinatorial Chemistry & High Throughput Screening**, v. 9, p. 213-228, 2006. doi: 10.2174/138620706776055539.
- 36 VERMA, J.; KHEDKAR, V. M.; COUTINHO, E. C. 3D-QSAR in drug design - a review. **Current Topics in Medicinal Chemistry**, v. 10, n. 1, p. 95-115, 2010. doi: 10.2174/156802610790232260.
- 37 TROPSHA A. Best practices for QSAR model development; validation; and exploitation. **Molecular Informatics**, v. 29, n. 6-7, p. 476-488, 2010. doi: 10.1002/minf.201000061.

- 38 FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust but verify: on the importance of chemical structure curation in chemoinformatics and QSAR modeling research. **Journal of Chemical Information and Modeling**, v. 50, n. 7, p. 1189-1204, 2010. doi: 10.1021/ci100176x.
- 39 LEWIS, R. A.; WOOD, D. Modern 2D QSAR for drug discovery. **Wiley Interdisciplinary Reviews: computational molecular science**, v. 4, n. 6, p. 505-522, 2014. doi: 10.1002/wcms.1187.
- 40 HERITAGE, T. W.; LOWIS, D. R. Molecular Hologram QSAR. In: PARRILL, A. L.; REDDY, M. R. (Ed.). **Rational drug design: novel methodology and practical applications**. 1999, cap. 14, p. 212-225. doi: 10.1021/bk-1999-0719.ch014.
- 41 OPREA, T. I. 3D QSAR modeling in drug design. In: LANGENAEKER, W.; WINTER, H. DE; BULTINCK, P.; TOLLENAERE, J. P. (Ed.). **Computational medicinal chemistry for drug discovery**. CRC Press, 2003, cap. 22. doi: 10.1201/9780203913390.ch22.
- 42 LOWIS, D. R. HQSAR: a new; highly predictive QSAR technique. **Tripes Tech Notes**, v. 1, n. 5, p. 1-17, 1997.
- 43 SALUM, L. B.; ANDRICOPULO, A. D. Fragment-based QSAR strategies in drug design. **Expert Opinion on Drug Discovery**, v. 5, n. 5, p. 405-412, 2010. doi: 10.1517/17460441003782277.
- 44 SALUM, L. B.; ANDRICOPULO, A. D. Fragment-based QSAR: Perspectives in drug design. **Molecular Diversity**, v. 13, n. 3, p. 277-285, 2009. doi: 10.1007/s11030-009-9112-5.
- 45 CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, v. 110, n. 18, p. 5959-5967, 1988. doi: 10.1021/ja00226a005.
- 46 GOODFORD, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. **Journal of Medicinal Chemistry**, v. 28, n. 7, p. 849-857, 1985. doi: 10.1021/jm00145a002.
- 47 KUBINYI, H. Comparative molecular field analysis (CoMFA). In: **Handbook of Chemoinformatics: from data to knowledge**. 2003, cap. 44, p. 1555-1574. doi: 10.1002/9783527618279.ch44d.
- 48 KLEBE, G.; ABRAHAM, U.; MIETZNER, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. **Journal of Medicinal Chemistry**, v. 37, n. 4, p. 4130-4146, 1994. doi: 10.1021/jm00050a010.
- 49 BURROWS, J. N. et al. Designing the next generation of medicines for malaria control and eradication. **Malaria Journal**, v. 12, n. 187, p. 1-20, 2013. doi: 10.1186/1475-2875-12-187.
- 50 NJOGU, P. M. et al. Computer-aided drug discovery approaches against the tropical infectious diseases malaria; tuberculosis; trypanosomiasis; and leishmaniasis. **ACS Infectious Diseases**, v. 2, n. 1, p. 8-31. doi: 10.1021/acsinfecdis.5b00093.
- 51 WRIGHT, M. H. et al. Protein myristoylation in health and disease. **Journal of Chemical Biology**, v. 3, n. 1, p. 19-35, 2010. doi: 10.1007/s12154-009-0032-8.
- 52 FARAZI, T. A.; WAKSMAN, G.; GORDON, J. I. The biology and enzymology of protein N-myristoylation. **Journal of Chemical Biology**, v. 276, n. 43, p. 39501-39504, 2001. doi: 10.1074/jbc.R300008200.
- 53 BELL, A.S. et al. Selective inhibitors of protozoan protein N-myristoyltransferases as starting points for tropical disease medicinal chemistry programs. **PLOS Neglected Tropical Diseases**, v. 6, n. 4, 2012. doi: 10.1371/journal.pntd.0001625.
- 54 WRIGHT, M. H. et al. Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. **Nature Chemistry**, v. 6, n. 2, p. 112-121, 2014. doi: 10.1038/nchem.1830.
- 55 GONCALVES, V. et al. Discovery of Plasmodium vivax N-myristoyltransferase inhibitors: screening; synthesis; and structural characterization of their binding mode. **Journal of Medicinal Chemistry**, v. 55, n. 7, p. 3578-3582, 2012. doi: 10.1021/Jm300040p.

- 56 YU, Z.Y. et al. Design and synthesis of inhibitors of Plasmodium falciparum N-myristoyltransferase; a promising target for antimalarial drug discovery. **Journal of Medicinal Chemistry**, v. 55, n. 20, p. 8879-8890, 2012. doi: 10.1021/Jm301160h.
- 57 RACKHAM, M. D. et al. Discovery of novel and ligand-efficient inhibitors of Plasmodium falciparum and Plasmodium vivax N-myristoyltransferase. **Journal of Medicinal Chemistry**, v. 56, n. 1, p. 371-375, 2013. doi: 10.1021/jm301474t.
- 58 RACKHAM, M. D. et al. Design and synthesis of high affinity inhibitors of Plasmodium falciparum and Plasmodium vivax N-myristoyltransferases directed by ligand efficiency dependent lipophilicity (LELP). **Journal of Medicinal Chemistry**, v. 57, p. 2773-2788, 2014. doi: 10.1021/jm500066b.
- 59 CLARK, M.; CRAMER, R. D.; VAN OPDENBOSCH, N. Validation of the general purpose tripos 5.2 force field. **Journal of Computational Chemistry**, v. 10, n. 8, p. 982-1012, 1989. doi: 10.1002/jcc.540100804.
- 60 POWELL, M. J. D. Restart procedures for the conjugate gradient method. **Mathematical Programming**, v. 12, n. 1, p. 241-254, 1977. doi: 10.1007/BF01593790.
- 61 MADHAVAN T. et al. QSAR analysis on PfPK7 inhibitors using HQSAR; CoMFA; and CoMSIA. **Medicinal Chemistry Research**, v. 21, n. 6, p. 681-693, 2012. doi: 10.1007/s00044-011-9572-x.
- 62 TROSSINI, G. H. G. et al. Quantitative structure-activity relationships for a series of inhibitors of cruzain from Trypanosoma cruzi: molecular modeling, CoMFA and CoMSIA studies. **Journal of Molecular Graphics and Modelling**, v. 28, n. 1, p. 3-11, 2009. doi: 10.1016/j.jmgm.2009.03.001.
- 63 WEBER, A. et al. 3D QSAR selectivity analyses of carbonic anhydrase inhibitors: insights for the design of isozyme selective inhibitors. **Journal of Chemical Information and Modeling**, v. 46, n. 6, p. 2737-2760, 2006. doi: 10.1021/ci600298r.
- 64 KUMAR, S. P. The effect of various atomic partial charge schemes to elucidate consensus activity-correlating molecular regions: a test case of diverse QSAR models. **Journal of Biomolecular Structure & Dynamics**, p. 1-20, 2015. doi: 10.1080/07391102.2015.1044474.
- 65 CASTILHO, M. S.; GUIDO, R. V. C.; ANDRICOPULO, A. D. 2D Quantitative structure-activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. **Bioorganic & Medicinal Chemistry**, v. 15, n. 18, p. 6242-6252, 2007. doi: 10.1016/j.bmc.2007.06.021.
- 66 TOBIAS, R. D. **An introduction to partial least squares regression**. 1995. Disponível em: <http://hbanaszak.mjr.uw.edu.pl/TempTxt/Shorts/Tobias_PLS_Intro.pdf>. Acesso em: 23 jan. 2016.
- 67 GOLBRAIKH, A.; TROPSHA, A. Beware of q²! **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 269-276, 2002. doi: 10.1016/S1093-3263(01)00123-1.
- 68 GASTEIGER, J.; MARSILI, M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. **Tetrahedron**, v. 36, n. 22, p. 3219-3228, 1980. doi: 10.1016/0040-4020(80)80168-2.
- 69 HALGREN, T. A. Merck Molecular Force Field. I. Basis; form; scope; parameterization; and performance of MMFF94. **Journal of Computational Chemistry**, v. 17, n. 5-6, p. 490-519. doi: 10.1002/(sici)1096-987x(199604)17:5/6<490::aid-jcc1>3.0.co;2-p.
- 70 JAKALIAN, A. et al. Fast; efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. **Journal of Computational Chemistry**, v. 21, n. 2, p. 132-146, 2000. doi: 10.1002/(sici)1096-987x(20000130)21:2<132::aid-jcc5>3.0.co;2-p.
- 71 WANG, J. et al. Antechamber: an accessory software package for molecular mechanical calculations. **Journal of Computational Chemistry**, v. 25, n. 2, p. 1157-1174, 2005. doi: 10.1016/j.jmgm.2005.12.005.
- 72 WANG, J. et al. Automatic atom type and bond type perception in molecular mechanical calculations. **Journal of Molecular Graphics and Modelling**, v. 25, n. 2, p. 247-260, 2006. doi: 10.1016/j.jmgm.2005.12.005.
- 73 NEESE, F. The ORCA program system. **Wiley Interdisciplinary Reviews: computational molecular science**, v. 2, n. 1, p. 73-78, 2012. doi: 10.1002/wcms.81.

- 74 BRENNEMAN, C. M.; WIBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. **Journal of Computational Chemistry**, v. 11, n. 3, p. 361-373, 1990. doi: 10.1002/jcc.540110311.
- 75 MULLIKEN, R. S. Electronic population analysis on LCAO-MO molecular wave functions. I. **Journal of Chemical Physics**, v. 23, n.10, p. 1833-1840, 1955. doi: 10.1063/1.1740588.
- 76 FRANCL, M. M. et al. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. **Journal of Chemical Physics**, v. 77, n. 7, p. 3654-3665, 1982. doi: 10.1063/1.444267.
- 77 HEHRE, W. J.; DITCHFIELD, R.; POPLE, J. A. Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. **Journal of Chemical Physics**, v. 56, n. 5, p. 2257-2261, 1972. doi: 10.1063/1.1677527.
- 78 STEPHENS, P. J. et al. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. **Journal of Chemical Physics**, v. 98, n. 45, p. 11623-11627, 1994. doi: 10.1021/j100096a001.
- 79 GADHE, C. G. et al. Various partial charge schemes on 3D-QSAR models for P-gp inhibiting adamantyl derivatives. **Bulletin of the Korean Chemical Society**, v. 32, n. 5, p. 1604-1612, 2011. doi:10.5012/bkcs.2011.32.5.1604.
- 80 MITTAL, R. R. Partial charge calculation method affects CoMFA QSAR prediction accuracy. **Journal of Chemical Information and Modeling**, v. 49, n. 3, p. 704-709, 2009. doi: 10.1021/ci800390m.
- 81 KROEMER, R. T.; HECHT, P.; LIEDL, K. R. Different electrostatic descriptors in comparative molecular field analysis: a comparison of molecular electrostatic and coulomb potentials. **Journal of Computational Chemistry**, v. 17, n. 11, p. 1296-1308, 1996. doi: 10.1002/(sici)1096-987x(199608)17:11<1296::aid-jcc2>3.0.co;2-h.
- 82 WATTS, K. S. et al. ConfGen: a conformational search method for efficient generation of bioactive conformers. **Journal of Chemical Information and Modeling**, v. 50, n. 4, p. 534-546, 2010. doi: 10.1021/ci100015j.
- 83 JAIN, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. **Journal of Medicinal Chemistry**, v. 46, p. 499-511, 2003. doi: 10.1021/jm020406h.
- 84 FRIESNER, R. A. et al. Glide: a new approach for rapid; accurate docking and scoring. 1. method and assessment of docking accuracy. **Journal of Medicinal Chemistry**, v. 47, n. 7, p. 1739-1749, 2004. doi: 10.1021/jm0306430.
- 85 ESWAR, N. et al. Comparative protein structure modeling using Modeller. **Current Protocols in Bioinformatics**, v. 5, p. 1-, 2006. doi: 10.1002/0471250953.bi0506s15.
- 86 SHEN, M.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein Science**, v. 15, p. 2507-2524, 2006. doi: 10.1110/ps.062416606.
- 87 RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. **Journal of Molecular Biology**, v. 7, n. 1, p. 95-99, 1963. doi: 10.1016/S0022-2836(63)80023-6.
- 88 LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283-291, 1993. doi: 10.1107/S0021889892009944.
- 89 CHEN, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. **Acta Crystallographica D: biological crystallography**, v. 66, n. 1, p. 12-21, 2010. doi: 10.1107/S0907444909042073.
- 90 BRASIL. Ministério da Saúde. Portal da Saúde. **Mapa de risco por município brasileiro de infecção por malária no ano de 2015.** Disponível em: <http://portal.arquivos.saude.gov.br/images/pdf/2017/fevereiro/32/Mapa-de-risco_malaria_2015.pdf>. Acesso em: 30 mar. 2017.