



LETÍCIA SANTOS GARCIA

**ESTUDOS COMPUTACIONAIS DE POTENCIAIS
INIBIDORES DA ENZIMA N-MIRISTOILTRANSFERASE DE
Plasmodium falciparum E *Leishmania donovani***

**LAVRAS – MG
2017**

LETÍCIA SANTOS GARCIA

**ESTUDOS COMPUTACIONAIS DE POTENCIAIS INIBIDORES DA ENZIMA N-
MIRISTOILTRANSFERASE DE *Plasmodium falciparum* E *Leishmania donovani***

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de Concentração Química/Bioquímica, para obtenção do título de Doutor.

Profa. Dra. Elaine Fontes Ferreira da Cunha
Orientadora

**LAVRAS - MG
2017**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Garcia, Letícia Santos.

Estudos computacionais de potenciais inibidores da enzima N-Miristoiltransferase de *Plasmodium falciparum* e *Leishmania donovani* / Letícia Santos Garcia. - 2017.

106 p. : il.

Orientador(a): Elaine Fontes Ferreira da Cunha.

.

Tese (doutorado) - Universidade Federal de Lavras, 2017.

Bibliografia.

1. QSAR-4D. 2. N-Miristoiltransferase. 3. Inibidores. I. Cunha, Elaine Fontes Ferreira da. . II. Título.

LETÍCIA SANTOS GARCIA

**ESTUDOS COMPUTACIONAIS DE POTENCIAIS INIBIDORES DA ENZIMA
N-MIRISTOILTRANSFERASE DE *Plasmodium falciparum* E *Leishmania donovani***

**COMPUTATIONAL STUDIES OF POTENTIAL INHIBITORS OF ENZYME
N-MIRISTOILTRANSFERASE OF *Plasmodium falciparum* AND *Leishmania donovani***

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de Concentração Química/Bioquímica, para obtenção do título de Doutor.

APROVADA em 28 de março de 2017.

Dr. Cleber Nogueira Borges	UFLA
Dr. Marcus Vinícius Juliaci Rocha	UFF
Dra. Melissa Soares Caetano	UFOP
Dr. Teodorico de Castro Ramalho	UFLA

Profa. Dra. Elaine Fontes Ferreira da Cunha
Orientadora

**LAVRAS - MG
2017**

RESUMO GERAL

A malária e a leishmaniose são doenças negligenciadas que atingem milhares de pessoas em todo o mundo. A malária é uma doença causada por protozoários do gênero *Plasmodium*. Nos últimos anos têm-se verificado a resistência do parasita a diversos medicamentos, incluindo o fármaco de primeira escolha para o tratamento da doença. Já a leishmaniose é provocada por protozoários parasitas do gênero *Leishmania* e a quimioterapia para esta doença também é insatisfatória, apresentando diversos efeitos colaterais, alta toxicidade e custo elevado. A N-miristoilação de proteínas é catalisada pela N-miristoiltransferase (NMT), um alvo essencial para *Plasmodium falciparum* e *Leishmania donovani*. Com o objetivo de estudar novos inibidores da NMT, a análise quantitativa da relação estrutura-atividade na quarta dimensão foi empregada. No trabalho da malária foram utilizados 83 compostos e realizados dez alinhamentos. Este trabalho sugeriu uma boa correlação com os resultados experimentais além de permitir o planejamento de novos inibidores da NMT do *Plasmodium falciparum*. Já o trabalho da leishmaniose foi aplicado a uma série de 77 inibidores da NMT de *Leishmania donovani*. Em seguida, foram propostos três novos compostos utilizando os modelos gerados. Adicionalmente, foi realizado um estudo de ancoramento molecular para investigar as afinidades de ligação e os modos de interação entre os compostos propostos e o sítio ativo da NMT. Além disso, foi realizada uma avaliação de ADME (absorção, distribuição, metabolismo e excreção) dos potenciais inibidores que demonstraram propriedades farmacocinéticas satisfatórias.

Palavras-chave: Malária. *Plasmodium falciparum*. Leishmaniose. *Leishmania donovani*. Inibidores. QSAR-4D. N-Miristoiltransferase.

GENERAL ABSTRACT

Malaria and leishmaniasis are neglected diseases that afflict thousands of people worldwide. Malaria is a disease caused by protozoa of the genus *Plasmodium*. In recent years the resistance of the parasite to several drugs, including the drug of first choice for the treatment of the disease, has been verified. Leishmaniasis is caused by protozoan parasites of the genus *Leishmania* and chemotherapy for this disease is also unsatisfactory, presenting several side effects, high toxicity and high cost. N-myristoylation of proteins is catalyzed by N-myristoyltransferase (NMT), an essential target for *Plasmodium falciparum* and *Leishmania donovani*. In order to study new NMT inhibitors, quantitative analysis of the structure-activity relationship in the fourth dimension was used. In the work of malaria, 83 compounds were used and ten alignments were performed. This work suggested a good correlation with the experimental results, besides allowing the planning of new inhibitors of *Plasmodium falciparum* NMT. Already the work of leishmaniasis was applied to a series of 77 NMT inhibitors of *Leishmania donovani*. Then, three new compounds were proposed using the generated models. In addition, molecular docking was performed to investigate the binding affinities and interaction modes between the proposed compounds and the NMT active site. Furthermore, an ADME (absorption, distribution, metabolism and excretion) evaluation was performed and potential inhibitors demonstrated satisfactory pharmacokinetic properties.

Keywords: Malaria. *Plasmodium falciparum*. Leishmaniasis. *Leishmania donovani*. Inhibitors. 4D-QSAR. N-myristoyltransferase.

LISTA DE FIGURAS

Figura 1 - Ciclo de vida do <i>Plasmodium</i> sp.....	11
Figura 2 - Estrutura dos principais fármacos utilizados no tratamento da malária.....	14
Figura 3 - Ciclo de vida do parasita <i>Leishmania</i> sp.....	15
Figura 4 - Fêmea do mosquito <i>Lutzomyia longipalpis</i>	16
Figura 5 - Distribuição geográfica da leishmaniose visceral.....	17
Figura 6 - Estrutura dos principais fármacos utilizados no tratamento da leishmaniose..	19
Figura 7 - Miristoilação.....	20
Figura 8 - Representação esquemática das principais etapas para a geração de modelos pela metodologia de QSAR-4D.....	25

SUMÁRIO

	PRIMEIRA PARTE.....	8
1	INTRODUÇÃO	8
2	REFERENCIAL TEÓRICO.....	10
2.1	Malária	10
2.1.1	A quimioterapia da malária	12
2.2	Leishmaniose.....	14
2.2.1	A quimioterapia da leishmaniose.....	17
2.3	Acilação de proteínas e N-miristoilação.....	20
2.3.1	N-Miristoiltransferase	21
2.4	Relação quantitativa estrutura-atividade	22
2.4.1	QSAR-4D.....	23
2.4.2	Validações em QSAR	25
2.4.2.1	Validação interna.....	26
2.4.2.1.1	Coefficiente de determinação (r^2)	26
2.4.2.1.2	Validação cruzada por ‘leave-one-out’ (loo) (q^2)	26
2.4.2.1.3	Fator de desajuste de Friedman (LOF).....	27
2.4.2.1.4	Y-Randomization (R^2_r) e R^2_p	27
2.4.2.2	Validação externa	28
2.4.2.2.1	R^2_{Pred}	28
2.4.2.2.2	R^2_m	28
2.4.2.2.3	$\overline{r^2}_m$ e ΔR_m^2.....	29
3	CONSIDERAÇÕES FINAIS.....	30
	REFERÊNCIAS.....	31
	ANEXO A	36
	SEGUNDA PARTE - ARTIGOS.....	41
	ARTIGO 1 QSAR study of N-myristoyltransferase inhibitors of antimalarial agents.....	41
	ARTIGO 2 Construction of 4D-QSAR models for use in the design of novel N-miristoiltransferase inhibitors of <i>Leishmania donovani</i>.....	73

PRIMEIRA PARTE

1 INTRODUÇÃO

As doenças negligenciadas são aquelas que afetam populações pobres e, por isso pouco tem sido feito para encontrar medicamentos para as mesmas. Uma das principais doenças negligenciadas é a malária, que é uma doença historicamente descrita causada por protozoários do gênero *Plasmodium*. A doença é transmitida pela picada das fêmeas de mosquitos do gênero *Anopheles* infectadas. Dentre as espécies que infectam o homem, o *Plasmodium falciparum* é a espécie que mais leva os pacientes a óbito. A doença foi responsável por mais de 214 milhões de casos e 438.000 mortes em 2015, segundo a Organização Mundial da Saúde (WHO, 2015). Já tem sido relatado sinais de resistência à artemisinina, que é o fármaco de primeira escolha para a terapêutica da malária. Além da artemisinina, tem sido verificada a resistência a diversos outros medicamentos usados para tratar a doença, além dos diversos efeitos colaterais apresentados pelos mesmos. Desta forma, o estudo de novos inibidores torna-se fundamental para o tratamento da malária.

Outra doença negligenciada é a leishmaniose, que é desencadeada por espécies do gênero *Leishmania*, que são transmitidas através da picada da fêmea de insetos flebotomíneos contaminados. A doença manifesta-se em uma de três formas: cutânea, mucocutânea e visceral. Dentre as formas de leishmaniose, a leishmaniose visceral, ou calazar, é a mais grave. O calazar é o resultado da infecção por *Leishmania donovani* e é responsável por cerca de 200.000 a 400.000 novos casos em todo o mundo e se não tratada pode ser fatal em mais de 95% dos casos (WHO, 2016a). Embora exista diversos quimioterápicos para tratar a leishmaniose, estes apresentam diversos efeitos colaterais, alta toxicidade, alto custo, via de administração que leva ao abandono do tratamento e/ou elevado potencial teratogênico. Devido a esses diversos fatores é necessário a busca por novos fármacos para o tratamento da leishmaniose que sejam mais eficazes e apresentem um baixo custo.

A miristoilação N-terminal de proteínas é uma importante modificação que ocorre co- e pós-traducionalmente, e é capaz de modular a interação de proteínas, inclusive com a membrana. Além disso, está relacionada com diversas vias de sinalização celular. A enzima N-miristoiltransferase catalisa a transferência do grupo miristato a partir do miristoil-CoA a resíduos de glicina N-terminal (TATE et al., 2014). Esta enzima representa um alvo promissor para o desenvolvimento de novos medicamentos, já que ela é essencial para *Plasmodium falciparum* e para *Leishmania donovani*.

Métodos de química computacional exercem uma função considerável no processo de descoberta e desenvolvimento de novos medicamentos. Uma ferramenta bastante utilizada é o estudo da correlação entre a estrutura e a atividade de fármacos. Neste método, considera-se que existe uma relação entre as propriedades de uma molécula e sua estrutura química, e então, buscam-se estabelecer relações matemáticas simples para descrever e prever a atividade de um conjunto de inibidores semelhantes (ANDRADE et al., 2010). O método de QSAR-4D permite avaliar diferentes graus de liberdade conformacional por meio da etapa de simulação da dinâmica molecular (HOPFINGER, et al., 1997).

Uma validação minuciosa dos modelos de QSAR gerados é importante para o planejamento de novos inibidores. As validações internas são importantes para a calibração do modelo, enquanto que a validação externa permite obter maior confiabilidade nas previsões de novos inibidores.

Neste trabalho foram realizados estudos de QSAR-4D com o propósito de selecionar características importantes que um potente inibidor deva exibir. Pretendeu-se também propor candidatos a fármacos inibidores da NMT para o tratamento da malária e da leishmaniose. Com a finalidade de atingir os objetivos propostos, este trabalho está dividido em duas partes. A primeira expõe uma visão geral sobre os estudos publicados na literatura e a segunda parte apresenta os dois artigos resultantes deste trabalho.

2 REFERENCIAL TEÓRICO

2.1 Malária

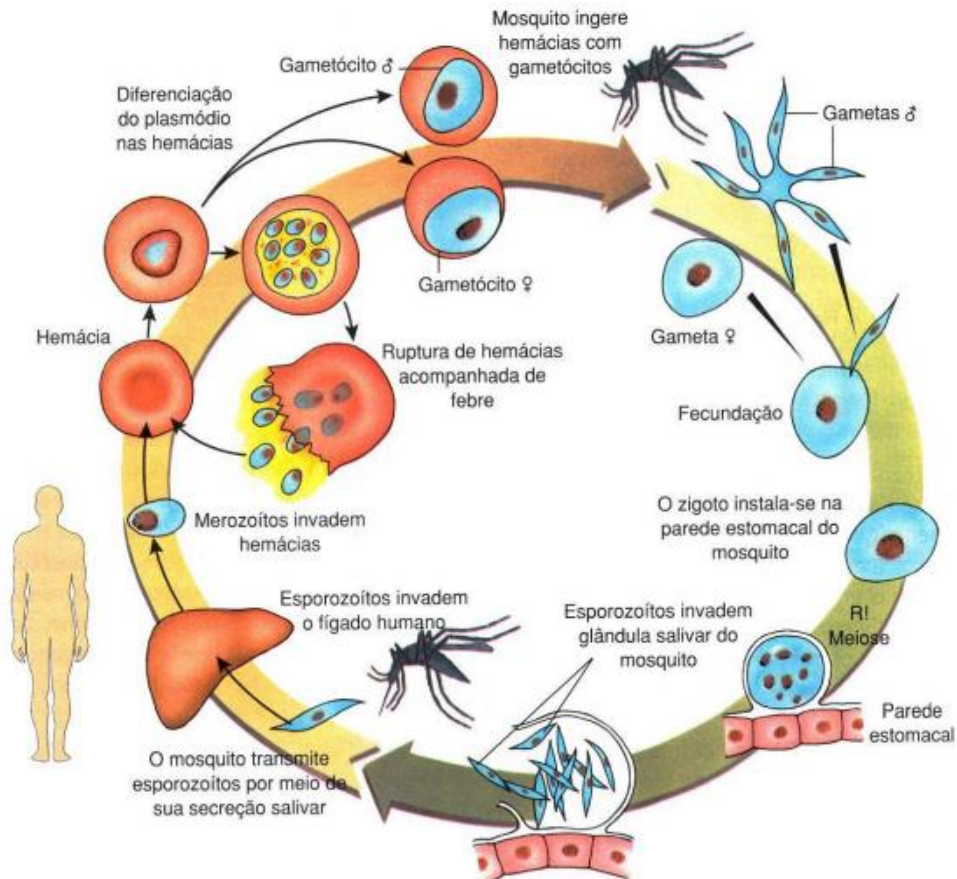
A malária, também conhecida como paludismo, foi inicialmente citada na era pré-Cristã por Hipócrates. Foi ele quem descreveu suas características de ocorrência sazonal e febre intermitente. Entretanto, foi somente no início do século XIX que o termo malária teve origem. Somente em 1880 que Charles Louis Alphonse Laveran conseguiu observar organismos em movimento ao observar, a fresco, o sangue de um paciente com a malária (NEVES, 2004; REY, 1991).

A doença é uma parasitose transmitida pela picada das fêmeas de mosquitos do gênero *Anopheles* infectadas. É causada por parasitas de cinco espécies do gênero *Plasmodium*: *P. falciparum*, *P. ovale*, *P. malariae*, *P. vivax* e *P. knowlesi*. Dentre estes, o *Plasmodium falciparum* é o mais prevalente e letal (NJOROGÉ et al., 2014).

O ciclo parasitário (Figura 1) inicia-se quando a fêmea do mosquito inocula juntamente com sua saliva as formas infectantes do *Plasmodium*, os esporozoítas. No fígado, os esporozoítas dão origem aos esquizontes teciduais e posteriormente a milhares de merozoítos que invadirão os eritrócitos. Depois de algumas gerações de merozoítos sanguíneos, ocorre a diferenciação em estágios sexuais, os gametócitos, que não mais se dividem e seguirão o desenvolvimento no mosquito vetor, dando origem aos esporozoítos (NEVES, 2004).

Durante o repasto sanguíneo, a fêmea do mosquito *Anopheles* ingere as formas sanguíneas do parasito, mas somente os gametócitos serão capazes de evoluir no inseto, formando o zigoto. Este atinge a parede do intestino médio, passando a ser chamado oocisto. Ocorre, então, a ruptura da parede do oocisto, sendo liberados os esporozoítos. Estes serão disseminados por todo o corpo do inseto através da hemolinfa, até atingir as células das glândulas salivares (NEVES, 2004; REY, 1991).

Figura 1 - Ciclo de vida do *Plasmodium* sp.



Fonte: SOUZA WEBER, 2010.

Apesar dos progressos na redução dos casos de malária e mortes, a doença foi responsável por mais de 214 milhões de casos e 438.000 mortes em 2015, segundo a OMS (WHO, 2015). Na região das Américas, estima-se que cerca de 112 milhões de pessoas em 21 países e territórios devem estar em risco de contrair a malária. O número de casos de malária confirmados na região caiu de 1,2 milhões em 2000 para 390.000 em 2014. Três países foram responsáveis por 77% dos casos em 2013: Brasil (37%), República Bolivariana da Venezuela (23%) e na Colômbia (17%). Entre 2000 e 2014, houve uma diminuição de mais de 75% na incidência da malária relatada em 15 dos 21 países, incluindo o Brasil. Entretanto, o Brasil ainda responde por quase metade das mortes devido à malária na região das Américas (WHO, 2015).

Intervenções globais para reduzir a prevalência e impacto da malária envolveram diversas estratégias, incluindo o controle de vetores, a quimiprolaxia e quimioterapia. Entretanto, um grande desafio para o tratamento da malária durante muitas décadas tem sido o surgimento de resistência do parasita às drogas disponíveis, necessitando, assim, de alterações periódicas nos

tratamentos utilizados (NJOROGE et al., 2014).

2.1.1 A quimioterapia da malária

A malária é uma doença antiga e referências na literatura antepassada e nas civilizações modernas relatam febres intermitentes e calafrios consistentes com os sintomas da malária. Em 2.700 a.C. o Cânon Chinês de medicina, o Nei Ching, já discutia sintomas da malária e as relações entre febres e melancolias prolongadas. O fisiologista grego Hipócrates foi o primeiro a observar a relação entre a proximidade de corpos d'água estagnados e a ocorrência de febres na população local. O termo malária surgiu exatamente desta relação entre a doença e os pântanos. A patologia passou a ser descrita como *ária cattiva* ou *mal'aria* (ar ruim) pelos italianos no século XIV. De forma semelhante, os franceses criaram o termo “paludismo”, cuja raiz significa pântano, para se referir à malária (FRANÇA; SANTOS; FIGUEROA-VILLAR, 2008).

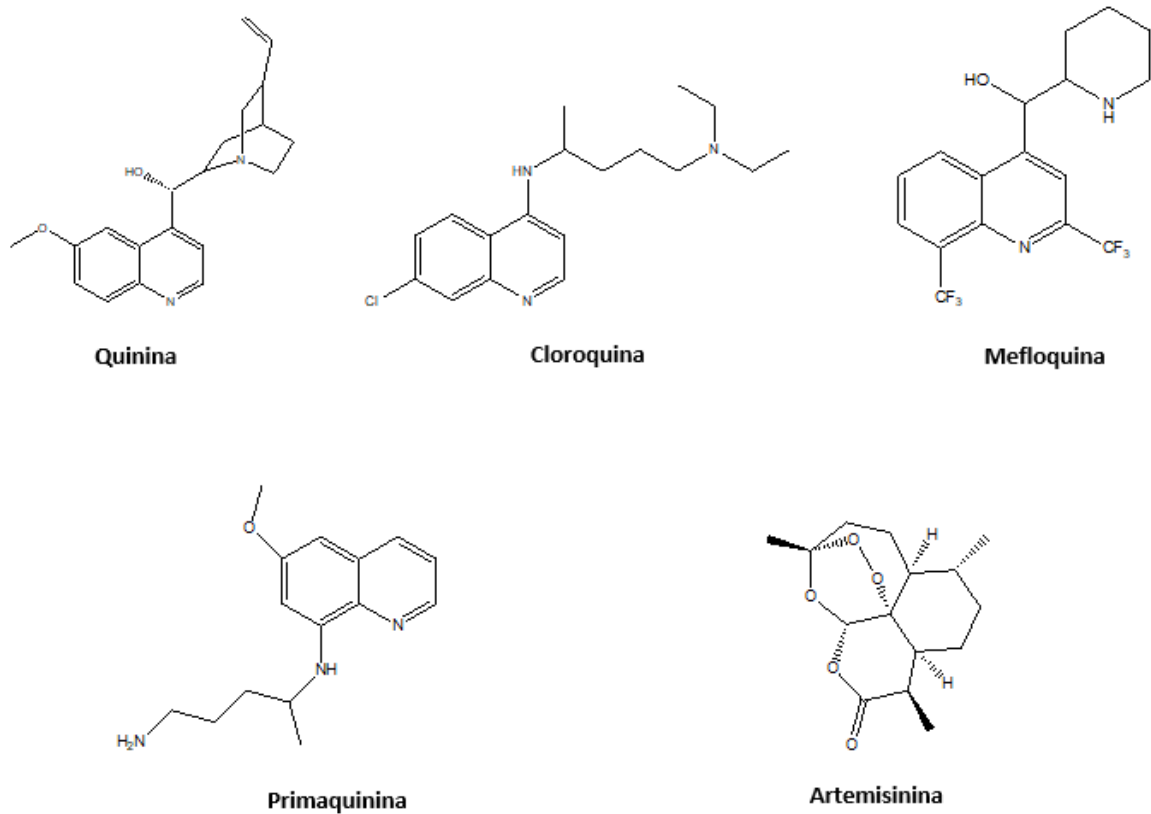
A quimioterapia da doença também é antiga e precedeu a descrição do parasita por cerca de 300 anos. No século XVII os padres jesuítas observaram a utilização, por populações indígenas da América do Sul, de chás e bebidas preparadas com a casca de uma árvore nativa do Peru, a *Cinchona sp.*, para o tratamento de alguns tipos de febre. Em 1820, Pierre Pelletier e Joseph Caventau, isolaram a quinina (Figura 2) da casca da casca da *Cinchona sp.*, (CUNICO et al., 2008) que foi o primeiro composto utilizado no tratamento da malária. Até a 2ª Guerra Mundial, a quinina era o único agente antiparasitário eficaz frente à malária. Este alcalóide quinolínico originou os fármacos antimaláricos cloroquina e mefloquina (Figura 2) (BARREIRO; RODRIGUES, 2001).

Durante a 2ª Guerra Mundial, as forças aliadas inauguraram um programa de pesquisa de antimaláricos, onde foram desenvolvidos vários fármacos, notavelmente a cloroquina (Figura 1) e a primaquina (SILVA; OLIVEIRA, 2005). A cloroquina, uma 4-aminoquinolina, tornou-se, então, o fármaco esquizonticida sanguíneo de escolha por mais de sessenta anos (TATE et al., 2014). Porém, o desenvolvimento de resistência dos parasitas da malária à cloroquina se deu por volta de 1950, (KUMAR; KUMARI; PANDEY, 2015). A mefloquina foi descoberta no Instituto Walter Reed do exército americano, nos EUA, para ser administrado em uma única dose diária (BARREIRO; RODRIGUES, 2001). A medicação é efetiva contra cepas de *Plasmodium falciparum* resistentes a cloroquina, embora sua resistência já tenha sido relatada principalmente no sudeste asiático, além de relatos de ocorrência na América do Sul, no leste da África e na Índia (CARVALHO, 2015).

A Primaquina (Figura 2) pertence à classe das 8-aminoquinolinas e foi sintetizada nos Estados Unidos em 1946, como parte dos esforços de tratamento e prevenção da malária nos campos de guerra. A primaquina é efetiva contra as formas exoeritrocitárias das espécies de *Plasmodium* capazes de infectar o homem. A droga nas doses terapêuticas empregadas é segura e bem tolerada. No entanto, a primaquina tem suas desvantagens: possui ação lenta, além de ser prejudicial para pessoas com deficiência na glicose-6-fosfato desidrogenase (G6PD) pois provoca reação adversa severa com risco de morte. O mecanismo pelo qual a primaquina e outros antimaláricos da família de 8-aminoquinolina causa a anemia hemolítica em pessoas deficientes em G6PD não é totalmente compreendido (UTHMAN et al., 2014). A G6PD faz parte da via das pentoses fosfato, que produz a coenzima nicotinamida adenina dinucleotídio fosfato reduzida (NADPH). NADPH está envolvida na produção de glutathione, que protege as células contra espécies reativas de oxigênio. Desta forma, os eritrócitos deficientes em G6PD são vulneráveis aos danos oxidativos e mais susceptíveis à hemólise devido à redução dos níveis de glutathione (CAPPELLINI; FIORELLI, 2008; KUMAR; KUMARI; PANDEY, 2015).

A artemisinina é um produto natural altamente eficaz contra a malária, isolado da *Artemisia annua* (TATE et al., 2014). Diferentemente das demais drogas existentes para essa finalidade, sua estrutura química (Figura 2) é um sesquiterpenlactona trioxano cuja ponte endoperóxido é essencial para sua atividade antimalárica (O'NEILL et al., 2010). Terapias de combinação de artemisinina (TCAs) têm sido cada vez mais utilizadas desde 2005, quando elas foram recomendados como tratamentos de primeira escolha para a malária pela Organização Mundial da Saúde (BONI; WHITE; BAIRD, 2016). Entretanto, a resistência à artemisinina e suas combinações estão sendo rotineiramente identificadas, como por exemplo, na Ásia, África, no Suriname, Guiana e Guiana Francesa. A resistência à artemisinina foi inicialmente identificada em estudos clínicos em 2006, embora uma análise retrospectiva de marcadores moleculares indicam que a resistência provavelmente surgiu em 2001 (WHO, 2016b). Desta forma, a descoberta de novos medicamentos que atuem por mecanismos distintos torna-se fundamental na redução das mortes associadas à malária por *Plasmodium falciparum*.

Figura 2 - Estrutura dos principais fármacos utilizados no tratamento da malária.



Fonte: Do autor (2017).

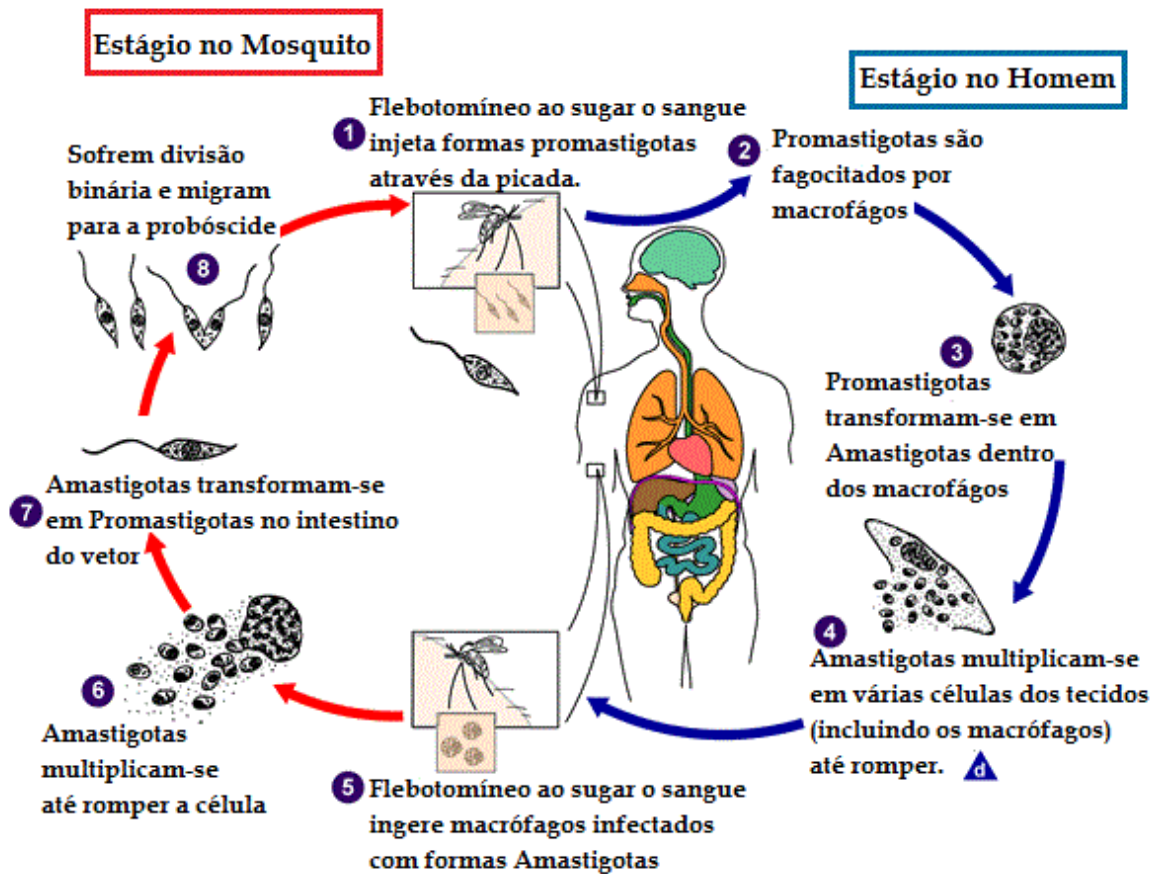
2.2 Leishmaniose

A leishmaniose é uma doença parasitária causada por espécies do gênero *Leishmania*, que são transmitidas de animais infectados para o homem através da picada da fêmea de insetos hematófagos conhecidos como flebotomíneos. Estes vetores são insetos pequenos de 2 a 4 mm de comprimento, de cor palha ou castanho claro, e conhecidos como mosquito palha ou birigui (NEVES, 2004; WHO, 2016c).

Os hospedeiros vertebrados são infectados quando formas promastigotas metacíclicas são inoculadas pelas fêmeas dos insetos vetores, durante o repasto sanguíneo (Figura 3). Nos hospedeiros mamíferos, os parasitas assumem a forma amastigota (aflageladas), arredondada e imóvel (3-6 μm), que se multiplicam obrigatoriamente dentro de células do sistema monocítico fagocitário (especialmente macrófagos). À medida que as formas amastigotas vão se multiplicando, os macrófagos se rompem, liberando parasitos que são fagocitados por outros macrófagos. O desenvolvimento no inseto vetor é iniciado quando fêmeas de flebotomíneos

ingerem sangue contendo macrófagos infectados com as formas amastigotas. Estas se transformam em promastigotas procíclicas no intestino médio do inseto que se diferenciam nas formas infectivas metacíclicas e migram, então, para a probóscide do inseto, onde poderão infectar um novo hospedeiro vertebrado (NEVES, 2004; SILVA, 2008).

Figura 3 - Ciclo de vida do parasita *Leishmania* sp.



Fonte: Adaptada de: CDC (Centers for Disease Control and Prevention)
<https://www.cdc.gov/dpdx/leishmaniasis/>, acessado em 25/01/2017.

Segundo a Organização Mundial da Saúde (OMS), a leishmaniose pode ser encontrada de três diferentes formas: visceral, cutânea e mucocutânea, sendo a forma visceral a mais grave (WHO, 2016a). A leishmaniose visceral, também conhecida como calazar, é uma infecção parasitária causada por *Leishmania donovani* e, se não for tratada, é fatal em mais de 95% dos casos dentro de 2 anos após o início da doença (WHO, 2016a). Dos dados disponíveis, a OMS estimou que 90% dos casos globais de LV ocorrem em 6 países, sendo que um deles é o Brasil.

De acordo com o Portal da Saúde (2014) do SUS (Sistema Único de Saúde), a *Lutzomyia longipalpis*, Figura 4, é a principal espécie transmissora da leishmaniose visceral (LV) no Brasil. Em posição de repouso, as asas destes insetos permanecem eretas e semiabertas, sendo

sua alimentação predominantemente noturna (PORTAL DA SAÚDE, 2014). O primeiro relato de LV no país foi feito em 1934, quando foram encontradas as formas amastigotas de *Leishmania* em cortes histológicos de fígado de pessoas que morreram com suspeita de febre amarela. Somente 20 anos depois é que se registrou o primeiro surto da doença em Sobral, no Ceará (GONTIJO; MELO, 2004).

Figura 4 – Fêmea do mosquito *Lutzomyia longipalpis*.



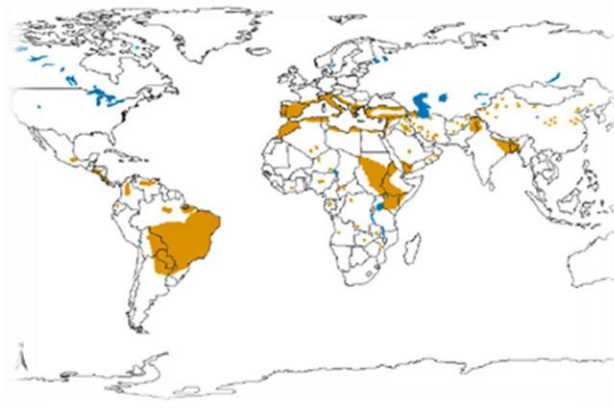
Fonte: http://www.raywilsonbirdphotography.co.uk/Galleries/Invertebrates/vectors/sand_fly.html. Acesso em 25 jan. 2017.

O Informe Epidemiológico das Américas de 2016 da OMS publicou que no período de 2001 a 2014 foram registrados um total de 48.720 casos de leishmaniose visceral na América Latina e média anual de 3.480 casos, sendo que 96,42% (46.976) estão concentrados no Brasil. A presença da leishmaniose está diretamente ligada à pobreza, mas fatores sociais, ambientais e climatológicos influenciam diretamente na epidemiologia da doença (WHO, 2016c). A LV é caracterizada por episódios irregulares de febre, perda de peso, aumento do baço e do fígado, e anemia (WHO, 2016c).

A leishmaniose visceral persiste como sério problema de saúde pública em várias partes do mundo. Em 2016, a OMS e o Gilead Sciences informaram que existem cerca de 200.000 a 400.000 novos casos em todo o mundo, sendo que a doença é endêmica em mais de 80 países (Figura 5) (WHO, 2016a).

A leishmaniose é classificada como uma doença negligenciada tropical. Por conseguinte, o interesse no desenvolvimento de novos compostos farmacêuticos é limitado porque seu tratamento precisa ser acessível para garantir o acesso da população pobre afetada. Além disso, os avanços no entendimento sobre a biologia das espécies de *Leishmania* não se traduziram em novos compostos quimioterápicos (MENEZES et al., 2015; CRUZ et al., 2009).

Figura 5 - Distribuição geográfica da leishmaniose visceral.



Fonte: http://www.who.int/leishmaniasis/leishmaniasis_maps/en/index.html. Acesso em: 26 jan. 2017.

2.2.1 A quimioterapia da leishmaniose

O uso medicinal de compostos de antimônio para diversos fins terapêuticos já era conhecido há séculos antes da era cristã, mas somente em 1912, Gaspar de Oliveira Vianna observou que o tártaro emético, Figura 6, era eficaz na terapêutica da leishmaniose. Entretanto, devido a sua alta toxicidade, os antimoniais trivalentes foram sendo substituídos por compostos pentavalentes (RATH et al., 2003). Em 1920, Bramachari desenvolveu o primeiro composto à base de antimônio pentavalente, derivado uréico do ácido p-aminofenil estibínico. Já em 1936, Schmidt criou o gluconato de antimônio (V) sódico, conhecido comercialmente como Solustibosan (Bayer) ou Pentostam (Glaxo Wellcome), Figura 6 (RATH et al., 2003).

Durante a Segunda Guerra Mundial, surgiu na França um medicamento alternativo ao até então gluconato de antimônio (V) sódico, o antimoniato de N-metil glucamina, comercializado como Glucantime (Rhône-Poulenc-Rohrer). Desde então, o Pentostan e o Glucantime têm sido considerados as drogas de primeira escolha no tratamento da leishmaniose (PORTAL DA SAÚDE, 2014). Estes fármacos devem ser administrados por via intramuscular ou intravenosa. Os efeitos secundários do tratamento com antimoniais incluem náuseas, dor abdominal, mialgia, inflamação pancreática, arritmia cardíaca e hepatite. Além disso, a longa duração do tratamento parenteral (até 28 dias) e o surgimento de resistência significativa também limitam a utilidade dessas drogas (PHAM; LOISEAU; BARRATT, 2013).

Outros fármacos, tais como anfotericina B, pentamidina, miltefosina e paromomicina também são utilizados. Por exemplo, a anfotericina B (Figura 6) é sistematicamente utilizada em pacientes imunocomprometidos e em caso de resistência aos antimoniais (ALVAR;

CROFT; OLLIARO, 2006). Contudo, a anfotericina B apresenta uma solubilidade limitada em água e baixa permeabilidade na membrana devido ao seu peso molecular relativamente elevado. Como resultado, esta droga é mal absorvida pelo trato gastrointestinal (AMIDON et al., 1995; LOURIA et al., 1958; PHAM; LOISEAU; BARRATT, 2013). Deste modo, a anfotericina B deve ser administrada por via intravenosa e frequentemente causa rigor, calafrios, e febre, associada a miocardite e nefrotoxicidade. As formulações de anfotericina B desenvolvidas para melhorar a farmacocinética e a biodisponibilidade têm diversas limitações como: muito dispendiosas e instáveis a temperaturas mais elevadas (MENEZES et al., 2015).

Apenas a miltefosina (Figura 6) pode ser administrada pela via oral, o que significou um grande avanço terapêutico no tratamento da LV. Esta via seria altamente preferível para tratamento em massa nas partes do mundo onde a LV é endêmica (CROFT et al., 1987). Originalmente desenvolvido como um fármaco anticancerígeno, demonstrou-se que a miltefosina possuía atividade antileishmania (CROFT et al., 1987). Entretanto, este fármaco não é desprovido de efeitos colaterais, os quais incluem distúrbios gastrointestinais, toxicidade renal (SUNDAR; OLLIARO, 2007) e um elevado potencial teratogênico (GHOSH; KAR; BERA, 2016).

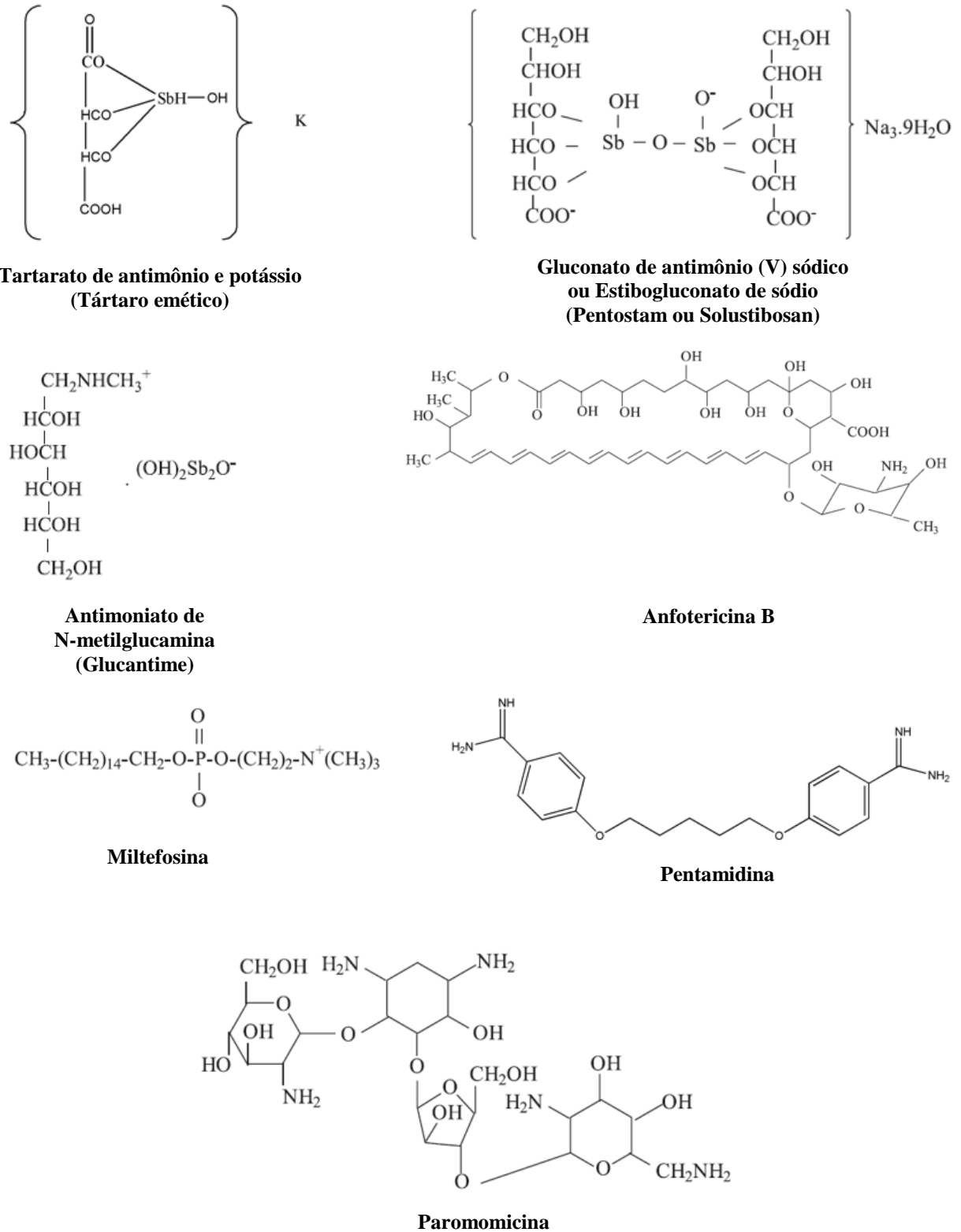
A pentamidina, Figura 6, foi inicialmente utilizada no tratamento de casos de calazar resistentes aos antimoniais. O tratamento com pentamidina apresenta um alto custo e a alta toxicidade desta droga também é fator limitante para o uso. Hipoglicemia, diabetes mellitus, hipotensão, alterações cardiológicas, nefrotoxicidade e, até mesmo, morte repentina já foram relatadas. Atualmente, a pentamidina é recomendada para ser utilizada em protocolos terapêuticos combinados (RATH, et al., 2003; MENEZES et al., 2015; SUNDAR; CHATTERJEE, 2006).

Uma droga alternativa para o tratamento da leishmaniose é a paromomicina (Figura 6). É um antibiótico aminoglicosídico usado para tratar infecções intestinais tais como amebíase e criptosporidiose. A atividade antileishmania da paromomicina foi identificada pela primeira vez na década de 1960 (NO, 2016). A paromomicina apresentou uma taxa de cura de 94,6% em ensaios clínicos de Fase IV na Índia, e sua formulação parenteral é atualmente usada para tratar LV. No entanto, formulações parentéricas podem causar reações adversas graves, incluindo nefrotoxicidade e ototoxicidade e mais raramente hepatotoxicidade (MENEZES et al., 2015).

Não existe uma droga ideal para o tratamento da leishmaniose. Cada uma das drogas existentes no mercado apresenta suas próprias limitações, tais como efeitos colaterais graves que levam ao abandono do tratamento (ALVAR; CROFT; OLLIARO, 2006). Além disso, o alto custo da medicação também é uma barreira para o controle eficaz da doença em países

pobres. Dessa forma, faz-se necessário a busca por novos medicamentos eficazes, de baixo custo e com reduzidos efeitos colaterais.

Figura 6 - Estrutura dos principais fármacos utilizados no tratamento da leishmaniose.



Fonte: Do autor (2017).

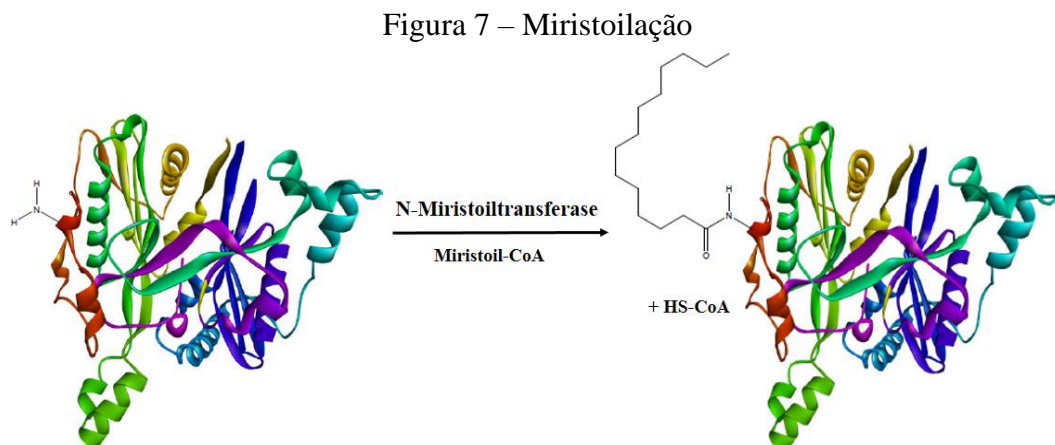
2.3 Acilação de proteínas e N-Miristoilação

A acilação é a adição covalente de ácidos graxos a proteínas. Os ácidos graxos mais comuns a serem adicionados a proteínas são o miristato e o palmitato, os quais são ácidos graxos saturados de 14 e 16 carbonos, respectivamente. A acilação é essencial em diversos processos biológicos importantes, incluindo o direcionamento de vesículas, a estabilização de interações proteína-membrana, além da sinalização celular (GOLDSTON et al., 2014).

A Miristoilação N-terminal (MIR), Figura 7, é uma das modificações proteína-lípido mais críticas; consiste na adição de um ácido graxo saturado de 14 carbonos, o miristato, a um resíduo de glicina N-terminal de uma determinada sequência proteica. O resíduo de glicina N-terminal é exposto, geralmente, como uma consequência do processo de excisão da metionina N-terminal durante a biossíntese de proteínas (GIGLIONE; FIEULAINÉ; MEINNEL, 2015). É um processo irreversível e geralmente ocorre co-traducionalmente, embora possa ocorrer pós-traducionalmente, quando o processamento proteolítico gera uma nova glicina N-terminal (GOLDSTON et al., 2014).

A MIR ocorre em aproximadamente 0,5% das proteínas eucarióticas e é realizada pela enzima N-miristoiltransferase (NMT). N-miristoilação pela NMT ocorre através do mecanismo de Bi-Bi, onde MYR-CoA liga-se a apo-enzima, induzindo uma alteração conformacional que permite que o substrato se ligue (RUDNICK et al., 1991; TATE et al., 2014).

MIR medeia interações proteína-proteína entre as proteínas de membrana que, por sua vez, facilita uma variedade de vias de transdução de sinal. Além disso, a miristoilação pode estar envolvida na estabilidade das proteínas e na associação de proteínas com membranas (TATE et al., 2014). A MIR é importante para a regulação e a sobrevivência celular e, assim, a NMT torna-se um alvo chave no desenvolvimento de novos quimioterápicos.



Fonte: Do autor (2017).

2.3.1 N-Miristoiltransferase

A N-Miristoiltransferase é uma enzima citosólica que catalisa a transferência do grupo miristato a partir da miristoil coenzima A (CoA) a resíduos de glicina amino terminal, em enzimas específicas (CURI et al., 2001; GOLDSTON et al., 2014; WRIGHT et al., 2014). Inicialmente o miristoil-CoA induz um rearranjo conformacional na NMT que revela seu sítio ativo. Em seguida, o miristato é transferido por uma reação de adição-eliminação nucleofílica para a glicina N-terminal do peptídeo, formando uma ligação amida que é seguida pela liberação sequencial do CoA e da proteína miristoilada (BRAND et al., 2011).

Como a miristoilação geralmente ocorre durante a tradução, enquanto a molécula nascente ainda está presa ao ribossomo, é necessária a prévia remoção da metionina inicial. A NMT não apresenta atividade intrínseca de aminopeptidase, sendo assim, é necessário a remoção do aminoácido metionina por uma aminopeptidase, de forma que a glicina na posição 2 torna-se o aminoácido terminal. Porém não são todas as glicinas N-terminais que serão miristoiladas. Para isto, elas devem possuir a sequência consenso Met-Gly-X-X-X-Ser/Thr-X-X (CURI et al., 2001; GOLDSTON et al., 2014).

A NMT é uma enzima tipicamente monomérica de 50 KDa que parece ser ubíqua em eucariotos. Em leveduras e protozoários a enzima está presente em uma única isoforma, enquanto em vertebrados e plantas existe 2 isoformas, a NMT1 e a NMT2 (GOLDSTON et al., 2014; SYLKE et al., 2016).

A enzima NMT de *Plasmodium falciparum* possui 410 aminoácidos, e a NMT de *Leishmania donovani*, 421 aminoácidos. As sequências primárias de aminoácidos dessas enzimas foram alinhadas com a sequência primária das duas isoformas da NMT humana (NMT1 com 496 aminoácidos e NMT2 com 410 aminoácidos) utilizando o servidor BLAST (ANEXO A) (ALTSCHUL et al., 1990). O alinhamento entre a enzima NMT1 humana e a NMT de *Plasmodium falciparum* demonstrou uma identidade de 51%, e entre a NMT2 humana e a de *Plasmodium falciparum* também apresentou identidade de 51%. Já para o alinhamento entre a enzima NMT1 humana e a NMT de *Leishmania donovani* foi verificada uma identidade de 42%, e entre a NMT2 humana e a de *Leishmania donovani* foi de 44%. Finalmente, o alinhamento entre as enzimas de *Plasmodium falciparum* e a de *Leishmania donovani* apresentou uma identidade de 40%.

A inibição da NMT demonstrou ser essencial para a sobrevivência de diversos organismos parasitas (GOLDSTON et al., 2014) e, portanto, representa um alvo racional para o desenvolvimento de fármacos para o tratamento da malária e da leishmaniose. Uma ferramenta

útil para desenvolver novos compostos protótipos para a inibição da NMT é por meio do estudo de relações quantitativas entre a estrutura química e atividade biológica.

2.4 Relação quantitativa estrutura-atividade

O princípio elementar da relação quantitativa estrutura atividade (QSAR, do inglês *Quantitative Structure-Activity Relationship*) é que as propriedades de uma molécula são determinadas por sua estrutura molecular (SUTHERLAND; O'BRIEN; WEAVER, 2004). Portanto, a atividade biológica de um fármaco está relacionada com as características moleculares específicas, que são denominadas de descritores (ANDRADE et al., 2010).

Os estudos de QSAR têm amadurecido em uma ferramenta amplamente utilizada, que contribui substancialmente para o processo de descoberta de drogas (LILL, 2007). Essa metodologia economiza recursos e agiliza o processo de desenvolvimento de novas moléculas (KHAN et al., 2009).

Modelos de QSAR devem ser capazes de prever a atividade de uma série congênere. Além disso, eles são utilizados para fazer previsões sobre novos derivados a fármacos (SUTHERLAND; O'BRIEN; WEAVER, 2004). Uma vez que a estrutura é conhecida, qualquer descritor molecular pode ser calculado, sendo o composto de interesse já sintetizado ou não (YANMAZ et al., 2011).

Os primeiros estudos QSAR basearam-se em uma única propriedade físico-química, como a solubilidade para explicar o efeito biológico da molécula (QSAR 1D). Hansch e Fujita, consideraram que a atividade biológica observada é resultado da contribuição de diversas propriedades físico-químicas (QSAR 2D), e seu modelo tenta identificar a contribuição de cada fator no estabelecimento da resposta biológica (TAVARES, 2004). Informações significativas sobre um conjunto de dados podem ser extraídas, utilizando o QSAR 2D, entretanto, esses descritores não possuem informações relacionadas com as características tridimensionais dos compostos (OPREA, 2002).

No QSAR 3D, considera-se a estrutura tridimensional do ligante, e este modelo fornece uma informação estrutural completa, tendo em conta a topologia, composição e forma espacial da molécula em uma única conformação (KUZ'MIN; ARTEMENKO; MURATOV, 2008). Em 1988, o método de análise comparativa de campo molecular (CoMFA) foi publicado por Cramer et al. (1988). Este método, que é baseado em campo molecular, constituiu o primeiro verdadeiro método de QSAR 3D (KUBINYI, 1997).

Já o QSAR 4D foi proposto por Hopfinger e colaboradores (1997) e examina o espaço

conformacional dos ligantes. Este modelo é semelhante ao modelo 3D, porém, a quarta dimensão na metodologia de QSAR-4D corresponde à amostragem conformacional em função do tempo de simulação pela dinâmica molecular, permitindo assim, identificar a conformação que maximiza o valor da atividade em modelos de QSAR-3D (HOPFINGER et al., 1997; YANMAZ et al., 2011). Assim, após a obtenção de um conjunto de conformações geradas pela simulação da dinâmica molecular, estas devem ser utilizadas para gerar os descritores na metodologia de QSAR-4D.

Na ausência da estrutura do receptor biológico, diferentes protocolos são apresentados no QSAR 5D, como uma solução possível para o problema, permitindo a avaliação simultânea de diferentes cenários de ajuste induzido (VEDANI et al., 2005). No QSAR 6D, a estimativa simultânea de diferentes modelos de solvatação pode ser considerada (VEDANI et al., 2007).

Finalmente, o QSAR-7D inclui receptores reais ou modelos de receptores virtuais baseados em alvo como, por exemplo, os obtidos por modelagem de proteínas por homologia (ROY; KAR; NARAYAN, 2015; POLANSKI, 2009). Neste trabalho, será utilizada a metodologia de QSAR-4D no intuito de identificar características importantes que um potente inibidor da NMT deva apresentar.

2.4.1 QSAR-4D

O método de QSAR-4D foi desenvolvido por Hopfinger e colaboradores e permite explorar diferentes graus de liberdade conformacional e distintos alinhamentos na busca de uma conformação representativa do modo de interação dos farmacóforos (HOPFINGER et al., 1997; VENKATARAMAN; HOPFINGER, 1999).

O método envolve o cálculo de simulação de dinâmica molecular (SDM) para cada molécula e, assim, são gerados um enorme número de confôrmeros (BAK; POLANSKI, 2006). Estes confôrmeros são utilizados para obter um perfil de amostragem conformacional (CEP, do inglês *Conformational Ensemble Profile*). Assim, o diferencial deste método é que ele incorpora a flexibilidade conformacional do ligante.

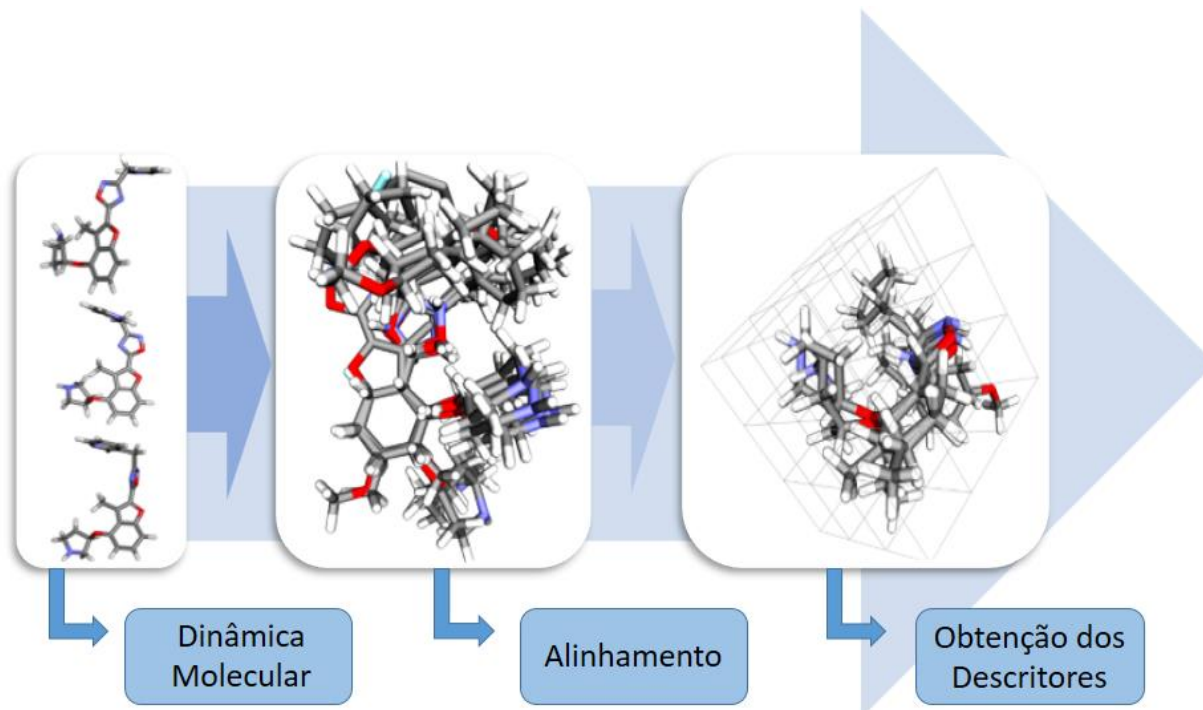
Cada conformação do CEP de cada molécula é então colocada em uma caixa tridimensional virtual, onde é definido o alinhamento (HOPFINGER; DUCA, 2000). O alinhamento é a maneira pela qual os compostos do conjunto de treinamento são comparados. A qualidade do modelo gerado depende do alinhamento testado, sendo assim, esta é considerada uma das etapas mais importantes na metodologia de QSAR-4D (CUNHA et al., 2005). Em geral, os alinhamentos são escolhidos de modo a abranger uma estrutura em comum dos compostos dos

grupos treinamento e teste. Alinhamentos usando átomos da direita, esquerda e do meio, ou que usam átomos que abrangem toda a região similar dos ligantes devem ser usados para garantir um bom alinhamento (OLIVEIRA; RAMALHO; CUNHA, 2009; ROMEIRO et al., 2005). A ordem dos átomos escolhidos para cada alinhamento é importante, já que durante o procedimento o primeiro átomo selecionado é posicionado na origem dos três eixos cartesianos e a molécula é girada de modo que o segundo átomo se posicione no eixo x, e o terceiro se posicione no eixo y (HOPFINGER et al., 1997).

O próximo passo é a seleção dos elementos de interação farmacofórica, IPEs (do inglês *Interaction Pharmacophore Elements*), que correspondem aos tipos de átomos que podem ocupar a célula da caixa. Os átomos de cada composto são classificados em sete tipos de IPEs, que correspondem aos tipos de interação que podem ocorrer no sítio ativo e estão relacionados com grupos farmacofóricos (ROMEIRO et al., 2005). Os IPEs descritos pelo programa de QSAR-4D são: (1) qualquer tipo de átomo, (2) interação não polar, (3) interação polar positiva, (4) interação polar negativa, (5) acceptor de ligação hidrogênio, (6) doador de ligação hidrogênio e (7) átomos em sistemas aromáticos (ANDRADE et al., 2010; HOPFINGER, 2001).

Na análise QSAR-4D, a frequência de ocupação de cada célula da caixa por cada IPE é registrada ao longo das conformações obtidas pela SDM e forma o conjunto de descritores de ocupação da célula (GCOD, do inglês *Grid Cell Occupancy Descriptors*) (VENKATARANGAN; HOPFINGER, 1999) que serão utilizados na construção dos modelos de QSAR. A Figura 8 mostra um esquema com as etapas para a obtenção de modelos de QSAR-4D.

Figura 8 - Representação esquemática das principais etapas para a geração de modelos pela metodologia de QSAR-4D.



Fonte: Do autor (2017).

2.4.2 Validações em QSAR

A validação dos modelos de QSAR desenvolvidos desempenha um papel substancial para a sua aplicação bem sucedida na previsão de novos compostos (ROY et al., 2013). As validações são divididas em validação interna e validação externa. A fim de realizar estas validações o conjunto de dados é dividido em um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento é composto pelas moléculas que serão utilizadas na construção dos modelos e participam, então, da validação interna dos modelos. Já o conjunto de teste não participa da construção dos modelos de QSAR e vários autores sugerem que a única forma de estimar o verdadeiro poder preditivo de um modelo é comparando as atividades previstas e observadas de um conjunto de teste suficientemente grande e que este não tenha sido usado no desenvolvimento do modelo (VEERASAMY et al., 2011). Assim, aproximadamente 20-30% do conjunto de dados é separado para compor o conjunto de teste. Os principais parâmetros utilizados nas validações interna e externa são descritos a seguir:

2.4.2.1 Validação interna

2.4.2.1.1 Coeficiente de determinação (r^2)

O coeficiente de determinação, r^2 ou R^2 , é uma medida do grau de ajuste da equação de regressão múltipla aos dados amostrais (Eq. 1).

$$R^2 = 1 - \frac{\sum_1^n (Y_{Exp} - Y_{Pred})^2}{\sum_1^n (Y_{Exp} - \bar{Y}_{Exp})^2} \quad (1)$$

Na equação acima, Y_{Exp} representa o valor da atividade experimental, enquanto que Y_{Pred} é a atividade predita e \bar{Y}_{Exp} é a média da atividade experimental. À medida que o valor de r^2 desvia-se 1, a qualidade de ajuste do modelo piora.

No entanto, deve-se notar que r^2 é simplesmente uma medida da qualidade do ajuste entre o modelo de cálculo e os valores experimentais, e não reflete o poder preditivo do modelo (GRAMATICA; PILUTTI, 2004). Também pode ser calculado o RMSEc (raiz quadrada do erro quadrático médio da calibração) que é uma medida da diferença entre os valores observados e preditos pelo modelo.

2.4.2.1.2 Validação cruzada por ‘leave-one-out’ (loo) (q^2)

Esta validação, que no português significa ‘deixe um de fora’, consiste em excluir um composto do conjunto de treinamento por vez. O modelo de QSAR é então reconstruído com base nas moléculas resultantes, e a atividade da molécula que foi excluída da construção do modelo é predita pela nova equação obtida. Este ciclo é repetido até que todas as moléculas do conjunto de treinamento tenham sido excluídas uma vez. Este parâmetro avalia a robustez do modelo e é considerado aceitável quando apresenta $q^2 > 0,5$ (Eq. 2).

$$q^2 = 1 - \frac{\sum_1^n (Y_{Exp} - Y_{Predcv})^2}{\sum_1^n (Y_{Exp} - \bar{Y}_{Exp})^2} \quad (2)$$

onde, Y_{Exp} indica o valor de atividade experimental, Y_{Predcv} o valor de atividade predita de acordo com a técnica *Leave-one-out*, e \bar{Y}_{Exp} o valor médio de atividade experimental, para os compostos do conjunto de treinamento (JAGIELLO et al., 2014; ROY et al., 2015).

Para permitir a comparação entre modelos com número diferente de variáveis os valores de q^2 devem ser transformados em $q^2_{ajustado}$, demonstrado pela Equação 3:

$$q^2_{ajustado} = \frac{1-(1-q^2)}{\frac{n-1}{n-p}} \quad (3)$$

Nesta equação q^2 corresponde ao quadrado do coeficiente de correlação após a validação cruzada, n corresponde ao número de compostos do conjunto de treinamento e p corresponde à quantidade de variáveis do modelo. Ainda pode ser calculado o RMSEcv (raiz quadrada do erro quadrático médio da validação cruzada).

2.4.2.1.3 Fator de desajuste de Friedman (LOF)

Estima a qualidade de um modelo. Esta medida penaliza adequadamente a adição de termos à equação de tal forma a resistir ao excesso de ajuste. A falta de ajuste (LOF) (Eq. 4) é usada para estimar a qualidade de um modelo, de modo que o melhor modelo receba a melhor pontuação de aptidão.

$$LOF = \frac{LSE}{\left\{1 - \frac{(c+dp)}{m}\right\}^2} \quad (4)$$

onde LSE é o erro de mínimos quadrados (calculado a partir da diferença entre os valores reais e calculados para o índice de atividade sobre o conjunto de dados), c é o número de funções de base no modelo, d é o fator de suavização, p é o número total das variáveis contidas em todas as funções de base, e m é o número de amostras (compostos) no conjunto de treinamento.

2.4.2.1.4 Y-Randomization (R^2_r) e R^2_p

A validação por *Y-Randomization* é realizada de modo a garantir a robustez do modelo de QSAR desenvolvido. O teste consiste em permutar os valores das atividades biológicas experimentais (variável dependente Y) mantendo fixa a matriz de descritores. O propósito do *Y-Randomization* é garantir que a correlação existente entre as variáveis X e Y não ocorre ao acaso. Para isto, os modelos obtidos ao randomizar as atividades biológicas devem ser ruins e, assim, o valor de R^2_r deve ser bem menor que o valor de R^2 (KIRALJ; FERREIRA, 2009; ROY et al., 2015).

O parâmetro R_p^2 , Eq. 5, relatado por Roy, assegura que os modelos não são obtidos por acaso. O valor de R_p^2 deve ser superior a 0,5 para ser considerado como modelo aceitável (ROY, 2007; VEERASAMY et al., 2011).

$$R_p^2 = R^2 \times \sqrt{R^2 - R_f^2} \quad (5)$$

2.4.2.2 Validação externa

2.4.2.2.1 R_{Pred}^2

Vários autores têm sugerido que a única maneira de estimar o verdadeiro poder preditivo de um modelo de QSAR é comparar as atividades previstas e observadas de um conjunto de teste, ou seja, por meio de compostos que não foram utilizados no desenvolvimento do modelo (VEERASAMY et al., 2011).

A capacidade preditiva de um modelo pode ser confirmada pelo valor de externa R_{Pred}^2 , Eq. 6.

$$R_{Pred}^2 = 1 - \frac{\sum_1^n (Y_{Exp(teste)} - Y_{Pred(teste)})^2}{\sum_1^n (Y_{Exp(teste)} - \bar{Y}_{treinamento})^2} \quad (6)$$

onde, $Y_{Exp(teste)}$ e $Y_{Pred(teste)}$ indicam os valores de atividade experimental e predita do conjunto de teste, respectivamente; $\bar{Y}_{treinamento}$ é o valor médio da variável dependente para o conjunto de treinamento. Um valor de R_{Pred}^2 superior a 0,5 pode ser considerado como um bom indicador de preditibilidade externa (ROY et al., 2009). O RMSEp (raiz quadrada do erro quadrático médio da predição) também pode ser medido.

2.4.2.2.2 R_m^2

Tem-se verificado que o valor de R_{Pred}^2 pode não ser suficiente para indicar a predição externa de um modelo (ROY et al., 2009). Como o valor de R_{Pred}^2 é dependente da média do conjunto de treinamento, ele pode não refletir a verdadeira capacidade preditiva de um modelo para um novo conjunto de dados. Para refinar ainda mais a habilidade preditiva dos modelos de QSAR a medida de $r_m^2(teste)$ foi introduzida por Roy, Eq. 7 (ROY et al., 2013):

$$R_m^2 = r^2 \left(1 - \left| \sqrt{r^2 - r_0^2} \right| \right) \quad (7)$$

onde, r_0^2 é o coeficiente de correlação quadrático entre os valores observados e preditos do conjunto de teste com intercepto zero.

O coeficiente de correlação quadrático entre os valores observados (eixo Y) e preditos (eixo X) dos compostos com e sem intercepto são calculados para a determinação de r_m^2 . Para um modelo aceitável, o valor de r_m^2 (teste) deve ser maior que 0,5 (ROY et al., 2009, 2013).

Inicialmente, o conceito de r_m^2 foi aplicado apenas para a validação do conjunto de teste, mais ele também é passível de aplicação para o conjunto de treinamento quando se considera a correlação entre os valores observado e predito por leave-one-out (LOO), para o conjunto de treinamento. O r_m^2 (total) inclui predição para ambos, conjunto de teste e treinamento. Assim, esta estatística é baseada em um grande número de compostos e pode ser utilizada na seleção dos melhores modelos (VEERASAMY et al., 2011).

2.4.2.2.3 $\overline{r_m^2}$ e ΔR_m^2

Variações do r_m^2 , também proposta por Roy et al (2013) são os valores de $\overline{r_m^2}$ (Eq. 8) e ΔR_m^2 (Eq.9), que podem ser calculados para as validações internas e externas.

$$\overline{r_m^2} = (r_m^2 + r_m'^2)/2 \quad (8)$$

$$\Delta R_m^2 = (|r_m^2 - r_m'^2|) \quad (9)$$

onde, $r_m'^2 = r^2 \times (1 - \sqrt{r^2 - r_0^2})$.

Tem sido sugerido que para o modelo ser considerado preditivo o valor de $\overline{r_m^2}$ deve ser maior que 0,5 e para ΔR_m^2 menor que 0,2.

3 CONSIDERAÇÕES FINAIS

Neste trabalho foram propostos modelos de QSAR-4D para inibidores da NMT de *Plasmodium falciparum* e de *Leishmania donovani*. Os modelos gerados fornecem informações importantes que podem auxiliar no planejamento de fármacos. No trabalho de antimalaricais foram testados dois grupos de teste e dez alinhamentos. O melhor modelo foi obtido a partir do Alinhamento B3, e gerou uma equação com sete descritores. O Modelo B3 mostrou uma qualidade estatística e habilidades preditivas satisfatórias. A análise de QSAR-4D indicou um papel importante dos grupos de aceptores de hidrogênio e grupos aromáticos, o que permitiu propor cinco estruturas. As mesmas provaram ser mais ativas do que o Composto **81**, o mais ativo da série, além de terem suas propriedades farmacocinéticas avaliadas pela regra de Lipinski. Assim, estas moléculas podem ser consideradas promissores protótipos para o tratamento da malária.

No trabalho da leishmaniose foram construídos três grupos de teste e sete alinhamentos. O melhor modelo, Modelo A1, apresentou características importantes que podem ser aplicadas no desenvolvimento de novos inibidores da NMT. Com base nos descritores obtidos a partir do Modelo A1, foram propostos os compostos **A**, **B** e **C**, os quais demonstraram valores de atividade biológica superiores ao composto mais ativo desta série. Estes resultados foram corroborados por estudos de ancoramento molecular, que confirmam e ampliam os resultados do QSAR-4D, permitindo o melhor entendimento das interações entre os inibidores e o sítio ativo da NMT de *Leishmania donovani*. Além disso, os compostos **A**, **B** e **C** também passaram por uma avaliação de ADME e não demonstraram problemas de biodisponibilidade. Conseqüentemente, os compostos propostos podem ser propícios para o tratamento da leishmaniose. Os resultados obtidos pelos dois trabalhos demonstram-se passíveis de serem aplicados na busca por novos fármacos inibidores da NMT para o tratamento da malária e da leishmaniose.

REFERÊNCIAS

- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, New York, v. 215, n. 3, p. 403-410, Oct 1990.
- ALVAR, J.; CROFT, S.; OLLIARO, P. Chemotherapy in the treatment and control of leishmaniasis. **Advances in parasitology**, London, v. 61, p. 223-274, 2006.
- AMIDON, G. L. et al. A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. **Pharmaceutical research**, New York, v. 12, n. 3, p. 413-420, 1995.
- ANDRADE, C. H. et al. 4D-QSAR: Perspectives in Drug Design. **Molecules**, Basel, v. 15, n. 5, p. 3281-3294, May 2010.
- BAK, A.; POLANSKI, J. A 4D-QSAR study on anti-HIV HEPT analogues. **Bioorganic & Medicinal Chemistry**, Oxford, v. 14, n. 1, p. 273-279, Jan 2006.
- BARREIRO, E. J.; RODRIGUES, C. R. Sobre a química dos remédios, dos fármacos e dos medicamentos. **Cadernos Temáticos de Química Nova na Escola**, São Paulo, v. 3, p. 4-9, 2001.
- BONI, M. F.; WHITE, N. J.; BAIRD, J. K. The Community As the Patient in Malaria-Endemic Areas: Preempting Drug Resistance with Multiple First-Line Therapies. **Plos Medicine**, San Francisco, v. 13, n. 3, Mar 2016.
- BRAND, S. et al. Discovery of a novel class of orally active trypanocidal N-myristoyltransferase inhibitors. **Journal of medicinal chemistry**, Washington, v. 55, n. 1, p. 140-152, 2011.
- CAPPELLINI, M. D.; FIORELLI, G. Glucose-6-phosphate dehydrogenase deficiency. **Lancet**, London, v. 371, n. 9606, p. 64-74, Jan 2008.
- CARVALHO, A. G. **Dinâmica molecular aplicada à análise estrutural da hipoxantina-guanina-xantina fosforibosil transferase de *Plasmodium falciparum***. 2015. 85 p. Dissertação (Mestrado em Biologia Experimental) - Universidade Federal de Rondônia, Porto Velho, 2015.
- CRAMER, R. D. et al. Comparative molecular field analysis (CoMFA) 1: effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, Easton, v. 110, p. 5959-5967, 1988.
- CROFT, S. L. et al. The activity of alkyl phosphorylcholines and related derivatives against *Leishmania donovani*. **Biochemical pharmacology**, Oxford, v. 36, n. 16, p. 2633-2636, 1987.
- CRUZ, A. K. et al. Current treatment and drug discovery against *Leishmania* spp. and *Plasmodium* spp.: a review. **Current drug targets**, Hilversum, v. 10, n. 3, p. 178-192, 2009.
- CUNHA, E. F. F. da. et al. 4D-QSAR models of HOE/BAY-793 analogues as HIV-1 protease inhibitors. **Qsar & Combinatorial Science**, Braunschweig, v. 24, n. 2, p. 240-253, Mar 2005.
- CUNICO, WILSON et al. Fármacos antimalariais-história e perspectivas. **Rev. Bras. Farm**, Rio de Janeiro, v. 89, n. 1, p. 49-55, 2008.
- CURI, R. et al. **Entendendo a Gordura: Os Ácidos Graxos**. 1. ed. Barueri: Manole, 2001.

FRANÇA, T. C. C.; SANTOS, M. G. dos; FIGUEROA-VILLAR, J. D. Malária: aspectos históricos e quimioterapia. **Quim. Nova**, São Paulo, v. 31, n. 5, p. 1271-1278, 2008.

GHOSH, S.; KAR, N.; BERA, T. Oleanolic acid loaded poly lactic co-glycolic acid-vitamin E TPGS nanoparticles for the treatment of *Leishmania donovani* infected visceral leishmaniasis. **International Journal of Biological Macromolecules**, Guildford, v. 93, p. 961-970, 2016.

GIGLIONE, C.; FIEULAINÉ, S.; MEINNEL, T. N-terminal protein modifications: Bringing back into play the ribosome. **Biochimie**, Paris, v. 114, p. 134-146, Jul 2015.

GOLDSTON, A. M. et al. Acylation in trypanosomatids: an essential process and potential drug target. **Trends in Parasitology**, Oxford, v. 30, n. 7, p. 350-360, Jul 2014.

GONTIJO, C. M. F.; MELO, M. N. Leishmaniose visceral no Brasil: quadro atual, desafios e perspectivas. **Rev bras epidemiol**, São Paulo, p. 338-349, 2004.

GRAMATICA, P.; PILUTTI, P. Evaluation of different statistical approaches for the validation of quantitative structure–activity relationships. Italy, **ECVAM**, Ispra, v. 177, 2004.

HOPFINGER, A. J. et al. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. **Journal of the American Chemical Society**, Easton, v. 119, n. 43, p. 10509-10524, Oct 1997.

HOPFINGER, A. J. **4D-QSAR Package User's Manual, 3.0**. Lake Forest: The Chem21 Group, 2001.

HOPFINGER, A. J.; DUCA, J. S. Extraction of pharmacophore information from high-throughput screens. **Current Opinion in Biotechnology**, London, v. 11, n. 1, p. 97-103, Feb 2000.

JAGIELLO, K. et al. Direct QSPR: the most efficient way of predicting organic carbon/water partition coefficient (log K (OC)) for polyhalogenated POPs. **Structural Chemistry**, Budapest, v. 25, n. 3, p. 997-1004, Jun 2014.

KHAN, A. K. R. et al. Comparative QSTR study of saturated alcohols based on topological, constitutional, geometrical, and getaway descriptors. **Medicinal Chemistry Research**, Cambridge, v. 18, n. 9, p. 770-781, Dec. 2009.

KIRALJ, R.; FERREIRA, M. M. C. Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. **Journal of the Brazilian Chemical Society**, São Paulo, v. 20, n. 4, p. 770-787, 2009.

KUBINYI, H. QSAR and 3D QSAR in drug design .volume 1. methodology. **Drug Discovery Today**, Kidlington, v. 2, n. 11, p. 457-467, Nov. 1997.

KUMAR, S.; KUMARI, R.; PANDEY, R. New insight-guided approaches to detect, cure, prevent and eliminate malaria. **Protoplasma**, Wien, v. 252, n. 3, p. 717-753, May 2015.

KUZ'MIN, V. E.; ARTEMENKO, A. G.; MURATOV, E. N. Hierarchical QSAR technology based on the Simplex representation of molecular structure. **Journal of Computer-Aided Molecular Design**, Dordrecht, v. 22, n. 6/7, p. 403-421, June 2008.

LILL, M. A. Multi-dimensional QSAR in drug discovery. **Drug Discovery Today**, Kidlington, v. 12, n. 23/24, p. 1013-1017, Dec. 2007.

- LOURIA, D. B. et al. Some aspects of the absorption, distribution, and excretion of amphotericin B in man. **Antibiotic Med.** New York, v. 5, n. 5, p. 295-301, 1958.
- MENEZES, J. P. B. de. et al. Advances in development of new treatment for leishmaniasis. **BioMed research international**, Norfolk, v. 2015, 2015.
- NEVES, D. P. **Parasitologia Humana**. 11. ed. São Paulo: Atheneu, 2004.
- NJOROGE, M. et al. Recent Approaches to Chemical Discovery and Development Against Malaria and the Neglected Tropical Diseases Human African Trypanosomiasis and Schistosomiasis. **Chemical Reviews**, Washington, v. 114, n. 22, p. 11138-11163, Nov 2014.
- NO, J. H. Visceral leishmaniasis: Revisiting current treatments and approaches for future discoveries. **Acta tropica**, Basel, v. 155, p. 113-123, 2016.
- OLIVEIRA, A. A.; RAMALHO, T. C.; CUNHA, E. F. F. da. QSAR Study of Androstenedione Analogs as Aromatase Inhibitors. **Letters in Drug Design & Discovery**, Cambridge, v. 6, n. 8, p. 554-562, Dez 2009.
- OMS (Organização Mundial de Saúde). **Informe Epidemiológico das Américas**. 2016. Disponível em:
<http://www.paho.org/hq/index.php?option=com_topics&view=readall&cid=6722&Itemid=40754&lang=es> Acesso em: 05 jan. 2017.
- O'NEILL, P. M.; BARTON, V. E.; WARD, S. A. The Molecular Mechanism of Action of Artemisinin-The Debate Continues. **Molecules**, Basel, v. 15, n. 3, p. 1705-1721, Mar 2010.
- OPREA, T. I. On the information content of 2D and 3D descriptors for QSAR. **Journal of the Brazilian Chemical Society**, São Paulo, v. 13, n. 6, p. 811-815, Nov./Dec. 2002.
- PHAM, T. T. H.; LOISEAU, P. M.; BARRATT, G. Strategies for the design of orally bioavailable antileishmanial treatments. **International journal of pharmaceuticals**, Amsterdam, v. 454, n. 1, p. 539-552, 2013.
- POLANSKI, J. Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. **Current medicinal chemistry**, Schiphol, v. 16, n. 25, p. 3243-3257, 2009.
- PORTAL DA SAÚDE. **Manual de Vigilância e Controle da Leishmaniose Visceral**. 2014. Disponível em: < <http://portalsaude.saude.gov.br/index.php/portarias-esclarecimentos/726-o-ministerio/o-ministerio-principal/secretaria-svs/vigilancia-de-a-a-z/leishmaniose-visceral-lv>> Acesso em: 09 jan. 2017.
- RATH, S. et al. Antimoniais empregados no tratamento da leishmaniose: estado da arte. **Química Nova**, São Paulo, v. 26, n. 4, p. 550-555, 2003.
- REY, L. Parasitologia. **Bases da parasitologia médica**. Rio de Janeiro: Guanabara Koogan, 2002.
- ROMEIRO, N. C. et al. Construction of 4D-QSAR models for use in the design of novel p38-MAPK inhibitors. **Journal of Computer-Aided Molecular Design**, Dordrecht, v. 19, n. 6, p. 385-400, 2005.
- ROY, K. On some aspects of validation of predictive quantitative structure-activity relationship models. **Expert Opinion on Drug Discovery**, Boston, v. 2, n. 12, p. 1567-1577, 2007.

ROY, P. P. et al. On Two Novel Parameters for Validation of Predictive QSAR Models. **Molecules**, Basel, v. 14, n. 5, p. 1660-1701, May 2009.

ROY, K. et al. Some case studies on application of "r(m)(2)" metrics for judging quality of quantitative structure-activity relationship predictions: Emphasis on scaling of response data. **Journal of Computational Chemistry**, New York, v. 34, n. 12, p. 1071-1082, May 2013.

ROY, K. et al. **Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment**. Massachusetts: Academic press, 2015.

ROY, K.; KAR, S.; NARAYAN, R. das. Statistical Methods in QSAR/QSPR. In: **A Primer on QSAR/QSPR Modeling**. Kolkata: Springer international publishing, p. 37-59, 2015.

RUDNICK, D. A. et al. Kinetic and structural evidence for a sequential ordered bi bi mechanism of catalysis by *Saccharomyces-cerevisiae* myristoyl-coa - protein N-myristoyltransferase. **Journal of Biological Chemistry**, Baltimore, v. 266, n. 15, p. 9732-9739, May 25 1991.

SILVA, D. G. **Padronização do cultivo de amastigotas axênicos e intracelulares de Leishmania spp. e análise da atividade leishmanicida de chalconas**. 2008. 120 p. Dissertação (Mestrado em Biotecnologia) - Universidade de Santa Catarina, Florianópolis, 2008.

SILVA, T. H. A. da; OLIVEIRA, M. T. de. Estudo de Modelagem Molecular de Complexos Ferriprotoporfirina IX e Quinolinocarbinolaminas Antimaláricas: Proposta de um Farmacóforo. **Quim. Nova**, São Paulo, v. 28, n. 2, p. 244-249, 2005.

SOUZA WEBER, L. de. et al. Uma abordagem de relacionamento no ensino de biologia com as questões ambientais meio ambiente e os efeitos sobre o corpo humano. **Revista monografias ambientais**, Santa Maria, v. 1, n. 1, p. 96-107, 2010.

SUNDAR, S.; CHATTERJEE, M. Visceral leishmaniasis-current therapeutic modalities. **Indian Journal of Medical Research**, New Delhi, v. 123, n. 3, p. 345, 2006.

SUNDAR, S.; OLLIARO, P. L. Miltefosine in the treatment of leishmaniasis: clinical evidence for informed clinical risk management. **Therapeutics and clinical risk management**, Auckland, v. 3, n. 5, p. 733, 2007.

SUTHERLAND, J. J.; O'BRIEN, L. A.; WEAVER, D. F. A comparison of methods for modeling quantitative structure-activity relationships. **Journal of Medicinal Chemistry**, Washington, v. 47, n. 22, p. 5541-5554, Oct. 2004.

SYLKE, M. et al. **Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery**. Weinheim: J. Wiley, 2016.

TATE, E. W. et al. N- Myristoyltransferase as a potential drug target in malaria and leishmaniasis. **Parasitology**, London, v. 141, n. 1, p. 37-49, Jan 2014.

TAVARES, L. C. QSAR: a abordagem de Hansch. **Química Nova**, São Paulo, v. 27, n. 4, p. 631-639, jul./ago. 2004.

UTHMAN, O. A. et al. Safety of 8-aminoquinolines given to people with G6PD deficiency: protocol for systematic review of prospective studies. **Bmj Open**, London, v. 4, n. 5, p. e004664, 2014.

VEDANI, A. Novel ligands for the chemokine receptor-3 (CCR3): a receptor-modeling study based on 5D-QSAR. **Journal of Medicinal Chemistry**, Washington, v. 48, n. 5, p. 1515-1527, Mar. 2005.

VEDANI, A. et al. Predicting the toxic potential of drugs and chemicals in silico: a model for the peroxisome proliferator-activated receptor-gamma (PPAR gamma). **Toxicology Letters**, Amsterdam, v. 173, n. 1, p. 17-23, Aug. 2007.

VEERASAMY, R. et al. Validation of QSAR models-strategies and importance. **International Journal of Drug Design & Discovery**, Chur, v. 3, p. 511-519, 2011.

VENKATARANGAN, P.; HOPFINGER, A. J. Prediction of ligand-receptor binding free energy by 4D-QSAR analysis: Application to a set of glucose analogue inhibitors of glycogen phosphorylase. **Journal of Chemical Information and Computer Sciences**, Washington, v. 39, n. 6, p. 1141-1150, Nov-Dec 1999.

WHO (World Health Organization). **Artemisinin and artemisinin-based combination therapy resistance**. Geneva, 2016b. Disponível em: <<http://apps.who.int/iris/bitstream/10665/250294/1/WHO-HTM-GMP-2016.11-eng.pdf>>. Acesso em: 09 jan. 2017.

WHO (World Health Organization). **Epidemiological Report in the Americas**. Leishmaniasis Report, Geneva, 2016c. Disponível em: <http://www2.paho.org/hq/index.php?option=com_docman&task=doc_view&gid=21608&itemid&Itemid=270>. Acesso em: 07 fev. 2017.

WHO (World Health Organization). **WHO and Gilead Sciences extend collaboration against visceral leishmaniasis**. Geneva, 2016a. Disponível em: <http://www.who.int/neglected_diseases/news/WHO_and_Gilead_Sciences_extend_collaboration/en/>. Acesso em: 07 fev. 2017.

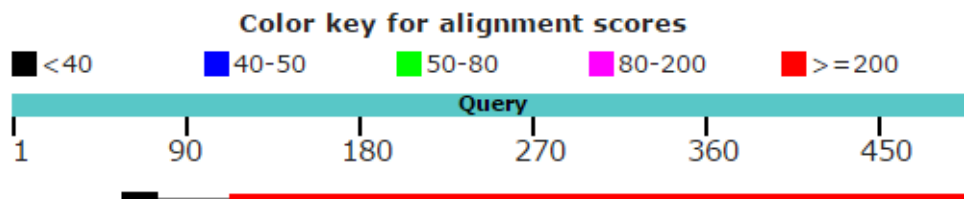
WHO (World Health Organization). **World Malaria Report 2015**. Geneva, 2015. Disponível em: <http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf> Acesso em: 05 jan. 2017.

WRIGHT, M. H. et al. Validation of N-myristoyltransferase as an antimalarial drug target using an integrated chemical biology approach. **Nature Chemistry**, London, v. 6, n. 2, p. 112-121, Feb 2014.

YANMAZ, E. et al. 4D-QSAR analysis and pharmacophore modeling: electron conformational-genetic algorithm approach for penicillins. **Bioorganic & Medicinal Chemistry**, Oxford, v. 19, n. 7, p. 2199-2210, Apr. 2011.

ANEXO A – Alinhamento entre as isoformas da enzima NMT humana e as enzimas NMT de *Plasmodium falciparum* e *Leishmania donovani*

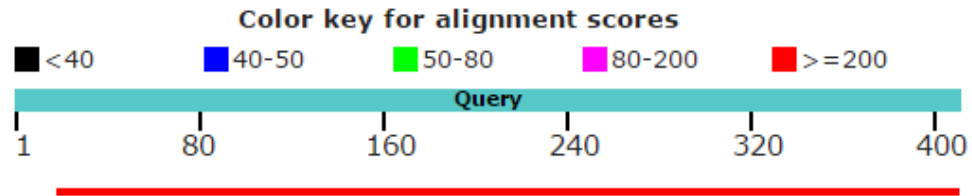
1 Alinhamento entre a isoforma NMT1 humana e a NMT de *Leishmania donovani*.



Max score	Total score	Query cover	E value	Ident	Accession
355	373	80%	2e-123	42%	Query_143895

1	MADESETAVKPPAPPLPQMMEGNGNGHEHCSDCENEEDNSYNRGGLSPANDTGAKKKKKKQKKKKEKGETDSAQDQPVK	80
1	MSR-----NPSNSDAA-----	11
81	MNSLPAERIQEIQKAIELFSVGQGPAKTMEEASKRSYQFWDTPVPKLG---EVVNTHGPV--EPDKDNIRQEPYTLPQG	155
12	-----HAFWSTQVPVQTEDETEKIVFAGPMDEPKTVADIPEEYPPIAST	55
156	FTWDALDLGDRGVLKELYTLLENYVEDDDNMFRFDYSPEFLWALRPPGWLQWHCGVRRVSSRKLVGFI SAIPANIHI	235
56	FEWTPNMEAADDIHAIYELLRDNYVEDDDSMFRFNYSSEFLQWALCPPSYIPDWHVAVRRKADKLLAFIAGVPVTLRM	135
236	-----YDEKKMVEINFLCVHKKLRSKRVPVLIREITRRVHLEGIFQAVYTAGVVLPKPVGTC	294
136	GTPKYMVKVKAQEKQEEAAKYDAPRHICEINFLCVHKQLREKRLAPILIKEVTRRVNRTNVWQAVYTAGVLLPTPYASG	215
295	RYWHRSLNPRKLI EVKFSHLSRNMT-----MQRTMKLYRLPETPKTAGLRPMETKDIPVHQLLTRYLKQFHLTPVMSQE	369
216	QYFHRSLNPEKLVEIRFSGIPAQYQKFQNP MAM LKRNYQLPNAPKNSGLREMKPSDVPQVRRILMNYLDNFDVGPVFSDA	295
370	EVEHWFYPQENIIDTFVVENANGEVTDFLSFYTL PSTIMNHPTHKSLKAAYSFYNVHTQTPLLDLMSDALVLAKMKGFDV	449
296	EISHYLLPRDGVVFTYVVEN-DKKVTDFFSFYRIPSTVIGNSNYNILNAAVHYAATSMPLHQLILDLLIVAHSRGF DV	374
450	FNALDLMENKTFLEKLFKFGIGDGNLQYYLYNWKCP SMGAEKVGLVLQ	496
375	CNMVEILDNRSFVEQLKFGAGDGH LRYFFYNWAYPKIKPSQVALVML	421

2 Alinhamento entre a isoforma NMT2 humana e a NMT de *Leishmania donovani*.



Max score	Total score	Query cover	E value	Ident	Accession
364	364	95%	3e-128	44%	Query_111755

```

1  MGSSHHHHHSSGLEVLLFQGPHMDEAAKHRYQFWDTQPVPKLDEV---ITSHGAIEPDK--DNVRQEPYSLPQGFMWDTL 75
1  -----MSRNPSNSDAA---HAFWSTQPVPQTEDETEKIVFAGPMDEPKTVADIPEEPYPIASTFEWWTP 61

76  DLSDAEVLKELYTLLNENYVEDDDNMFRDYSPEFLLWALRPPGWLLQWHCGVRSSNKLVGFISAIPANIRI----- 149
62  NMEAADDIHAIYELLRDNYVEDDSMFRFNYSEFLQWALCPPSYIPDWHVRRKADKLLAFIAGVPVTLRMGTPKYM 141

150 -----YDSVKKMVEINFLCVHKLRSKRVAPVLIREITRRVNLEGIFQAVYTAGVLPKPIATCRYWHRS 214
142 KVKAQEKGQEEEAKYDAPRHICEINFLCVHKLREKRLAPILIKEVTRRRVNRTNWQAVYTAGVLLPTPYASGQYFHRS 221

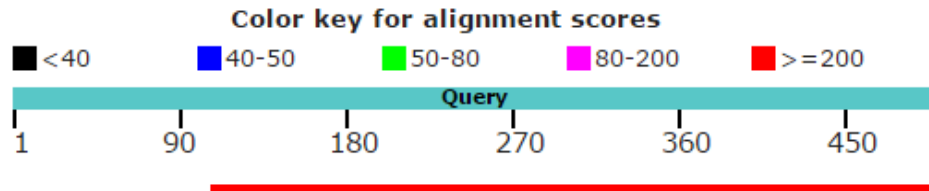
215 LNPKKLVEVKFSHLSRNMT-----LQRTMKLYRLPDVTKTSGLRPMEPKDIKSVRELINTYLKQFLAPVMDEEEVAHWF 289
222 LNPEKLVEIRFSGIPAQYQKFQNPMAMLKRNYQLPNAPKNSGLREMKPSDVPQVRRILMNYLDNFDVGPVFSDAEISHYL 301

290 LPREHIIDTFVESPNGLTDFLSFYTLPSTVMHHPAHKSLKAAYSFYNIHTETPLLDLMSDALILAKSKGFDVFNALDL 369
302 LPRDGVVFTYVEN-DKKVTDFFSFYRIPSTVIGNSNYILNAAYVHYYAATSMPLHQLILDLLIVAHSRGFDVCNMVEI 380

370 MENKTFLEKLKFGIGDGNLQYYLNWRCPGTDSEKVGLVLQ 410
381 LDNRSFVEQLKFGAGDGHLRYFYNWAYPKIKPSQVALVML 421

```

3 Alinhamento entre a isoforma NMT1 humana e a NMT de *Plasmodium falciparum*.



Max score	Total score	Query cover	E value	Ident	Accession
422	422	78%	8e-150	51%	Query_5929

```

1  MAESETAVKPPAPPLPQMMEGNGNGHEHCSDCENEEDNSYNRGGGLSPANDTGAKKKKKKQKKKKEKGETDSAQDQPVK  80
-----

81  MNSLPAERIQEIQKAIELFSVGQGPAKTMEEASKRSYQFWDTPVPKLGEV----VNTHGVPDPKDNIRQEPYTLPQGF  156
1  MNDDKKDFVGR-----DLYQL----IRNAKDIIKIDYKFWYTPVPKINDEFDENVNEPFISDNKVEDVRKEEYKLPSTY  71

157  TWDALDLGDRGVKELYTLLENYVEDDDNMFVFDYSPEFLWALRPPGWLQWHCGVRVSSRKLGVFISAIPANIIHY  236
72  AWCVCDITKENDRSDIYNLLTDNYVEDDDNVFRFNYSEFLWALSSPNYVKNWHIGVKYESTNKLGVFISAIPIDMCVN  151

237  DTEKKMVEINFLCVHKKLSKRVPVLIREITRRVHLEGIFQAVYTAGVVLPKPVGTCRYWHRSLNPRKLEVKFSHLSR  316
152  KNIIKMAEVNFLCVHKSLSKRRLAPVLIKEITRRINLESIQAIYTAGVYLPKPISTARYFHRSINVKKLEIGFSLNT  231

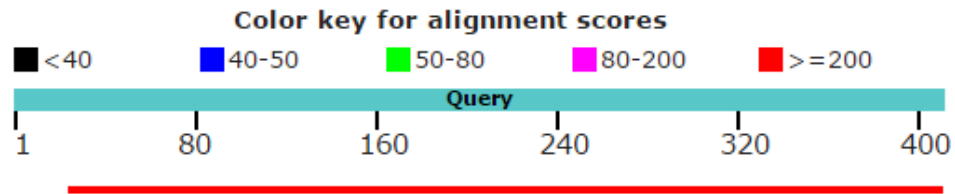
317  NMTMQRTMKLYRLPETPKTAGLRPMETKDIPVVHQLLTRYLKQFHLTPVMSQEEVEHWFYPQENIIDTFVVENANGEVTD  396
232  RLTMRAIKLYRIDDTLNIKNLRLMKKKDIDGLQKLLNEHLKQYNLHAIFSKEDVAHWFTPIDQVIYTYVNEE-NGEIKD  310

397  FLSFYTLPSTIMNHPHKSLSKAAYSFYNVHTQTPLLDLMSDALVLAKMKGFDFVFNALDLMENKTFLEKLFKFGIGDGNLQY  476
311  LISFYSLPSKVLGNKNYILNAAFSFYNIITTTTFFKNLIQDAICLAKRNNFDVFNALVMDNYSVFDLKFEGEGDGLKY  390

477  YLYNWKCPMSGAEKVLVLQ  496
391  YLYNWKCASCHPSKIGIVLL  410

```

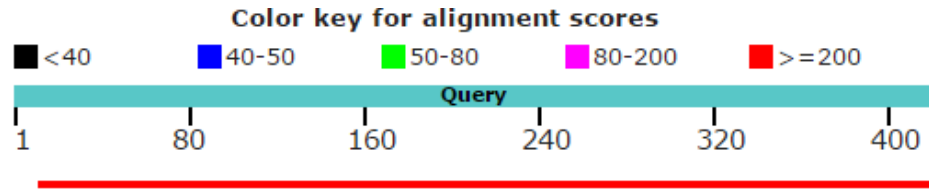
4 Alinhamento entre a isoforma NMT2 humana e a NMT de *Plasmodium falciparum*.



Max score	Total score	Query cover	E value	Ident	Accession
421	421	93%	7e-151	51%	Query_193191

1	MGSSHHHHHSSGLEVLFGPHMDEAAKHRYQFWDTPVPKLDVEI --- TSHGAIEPDK-DNVRQEPYSLPQGFMMWDTLD	76
1	MNDDKKDF --- VGRDLYQLIRNAKDKIKIDYKFWYTPVPKINDEFDENVNEPFISDNKVEDVRKEEYKLPSTGYAWCVCD	77
77	LSDAEVLKELYTLLENYVEDDDNMFRFDYSPEFLWLALRPPGWLQWHCGRVSSNKKLVGFISAIPANIRIYDSVKKM	156
78	ITKENDRSDIYNLLTDNYVEDDDNVFRFNYSSEFLWLALSSPNYVKNWHIGVKYESTNKLGVGFISAIPIDMCVNKNIKM	157
157	VEINFLCVHKKLRSKRVAPVLIREITRRVNLEGIFQAVYTAGVVLPKPIATCRYWHRSLNPKKLVKVFSHLSRNMTLQR	236
158	AEVNFLCVHKSLSKRLAPVLIKEITRRINLESINQAIYTAGVYLPKPISTARYFHRSINVKKLEIGFSLNTRLTMSR	237
237	TMKLYRLPDVTKTSGLRPMEPKDIKSVRELINTYLKQFHLAPVMDEEEVAHWFLPREHIIDTFVVEPNGKLTDFLSFYT	316
238	AIKLYRIDDTLNIKNLRLMKKKIDIGLQKLLNEHLKQYNLHAIFSKEDVAHWFTPIDQVIYTYVNEE-NGEIKDLISFYS	316
317	LPSTVMHHPAHKSLKAAYSFYNIHTETPLLDLMSDALILAKSKGDFVFNALDLMENKTFLEKLFKFGIGDGNLQYYLYNWR	396
317	LPSKVLGNKNYINLNAAFSFIYNTTTTTFKNLIQDAICLAKRNMDFVFNALVMDNYSVFDLKFGEEDGSLKYYLYNWK	396
397	CPGTDSEKVGVLVQ	410
397	CASCHPSKIGIVLL	410

5 Alinhamento entre as enzimas NMT de *Plasmodium falciparum* e de *Leishmania donovani*.



Max score	Total score	Query cover	E value	Ident	Accession
342	342	97%	2e-119	40%	Query_7681

1	MS-----RNPSNS-DAAHAFWSTQVPVQTEDETEKIVFAGPMDEPKTVADIPEEPYPIASTFEWTPNME	64
1	MNDKKDFVGRDLYQLIRNAKDKIKIDYKFWYTQVPVKINDEFENV-NEPFISDNKVEDVRKEEYKLPSPGYAWCVCDIT	79
65	AADDIHAIYELLRDNYVEDDDSMFRFNYSSEFLQWALCPPSYIPDWHVAVRRKADKKLLAFIAGVPVTLRMGTPKYMVKV	144
80	KENDRSDIYNLLTDNYVEDDDNVFRFNYSSEFLWALSSPNYVKNWHIGVKYESTNKLVGFIISAIPIDMCV-----	150
145	AQEKQEEEEAAKYDAPRHICEINFLCVHKQLREKRLAPILIKEVTRRVNRTNWNQAVYTAGVLLPTPYASGQYFHRSLNP	224
151	-----NKNIKMAEVNFLCVHKSLSKRLAPVLIKEITRRINLESIWQAIYTAGVYLPKPISTARYFHRISINV	218
225	EKLVEIRFSGIPAQYQKFNPMAMLRNYQLPNAPKNSGLREMKPSDVPQVRRILMNYLDNFDVGPVFSDAEISHYLLPR	304
219	KKLEIGFSCLNRLT-----MSRAIKLYRIDDTLNIKRLMKKKDIDGLQKLLNEHLKQYNLHAIFSKEDVAHWFTPI	293
305	DGVVFTYVENDKKVTDFFSFYRIPSTVIGNSNYNILNAAVHYAATSMPLHQLILDLLIVAHSRGFDVCMVEILDNR	384
294	DQVIYTYVNEENGEIKDLISFYSLPSKVLGNNKYNILNAAF SFYNIITTTTTFKNLIQDAICLAKRNNFDVFNALVMDNY	373
385	SFVEQLKFGAGDGLRYYFYNWAYPKIKPSQVALVML	421
374	SVFQDLKFGEGDGLKYYLYNWKCASCHPSKIGIVLL	410

SEGUNDA PARTE

ARTIGO 1

QSAR study of N-myristoyltransferase inhibitors of antimalarial agents.

Artigo redigido em concordância com as normas do periódico “*Interdisciplinary Sciences*”

(Versão preliminar)

Letícia Santos-Garcia¹, Tamiris Maria de Assis¹, Letícia Cristina de Assis¹, Daniela Rodrigues Silva¹, Giovanna Cardoso Gajo¹, Teodorico C. Ramalho^{1,2*}, Elaine F. F. da Cunha¹

¹Department of Chemistry, Federal University of Lavras, P.O. Box 3037, 37200-000, Lavras, MG, Brazil

²Center for Basic and Applied research, University Hradec Kralove, Hradec Kralove, Czech Republic.

Abstract Malaria is a disease caused by protozoan parasites of the genus *Plasmodium* that affects thousands of people worldwide. In recent years there have been parasite resistance to several drugs, including the first-line antimalarial treatment. With the aim of proposing new drugs candidates for the treatment of disease, 4D-QSAR methodology was applied to a series of 83 N-myristoyltransferase inhibitors, synthesized by Leatherbarrow et al. The QSAR models were developed using 63 compounds, the training set, and externally validated using 20 compounds, the test set. Ten different alignments for the two test sets were tested and the models were generated by the technique that combines genetic algorithms and partial least squares. The best model shows $r^2 = 0.757$, $q^2 = 0.702$, $q^2_{\text{adjusted}} = 0.634$, $R^2_{\text{pred}} = 0.746$, $R^2_{\text{m}} = 0.716$, $\Delta R^2_{\text{m}} = 0.133$, $R^2_{\text{p}} = 0.609$ and $R^2_{\text{r}} = 0.110$. This work suggested a good correlation with the experimental results and allows the design of new potent N-myristoyltransferase inhibitors.

Keywords: Malaria; *Plasmodium falciparum*; Inhibitors; N-myristoyltransferase; 4D-QSAR

1 Introduction

Malaria is a mosquito-borne protozoal infection caused by five parasites of five species of the genus *Plasmodium*: *P. falciparum*, *P. ovale*, *P. malariae*, *P. vivax*, and *P. knowlesi* [1]. Among these, *Plasmodium falciparum* is the most prevalent and lethal [2]. Malaria remains a major public health problem in many countries of the world despite progress in reducing malaria cases and deaths, malaria was responsible for over 214 million cases and 438,000 deaths in 2015 [3]. Researchers have identified the N-myristoyltransferase (NMT) enzyme as an important target for a generation of drugs to be used for the treatment of malaria. NMT catalyzes the transfer of the myristoyl group from a myristoyl coenzyme A (CoA) to the N-terminal glycine residue after the targeted protein[4]. N-terminal myristoylation (MYR) by NMT occurs through the Bi-Bi mechanism, where MYR-CoA binds to the apo-enzyme, inducing a conformational change that allows the NMT substrate to bind [5, 6]. *P. falciparum* has a single NMT isoform and mRNA is expressed in asexual blood-stage forms [7].

Over the past 50 years, the parasite resistance to chloroquine and sulphadoxine-pyrimethamine in endemic countries has been noted [8, 9]. In addition, signs of resistance to artemisinin-based combination therapies (ACTs) have been detected. ACTs are the first-line treatment for malaria and thus, new drugs are constantly required [6, 10-12].

MYR mediates protein-protein interactions between membrane proteins that in turn facilitate a variety of signal transduction pathways. Moreover, myristoylation can be involved in protein stability and association of proteins with membranes [6, 13]. The MYR is important for regulation and cell survival and thus, NMT becomes a key target in the development of new chemotherapeutics.

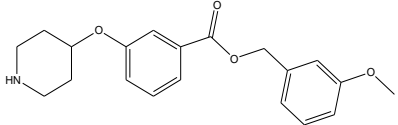
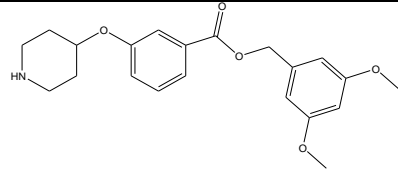
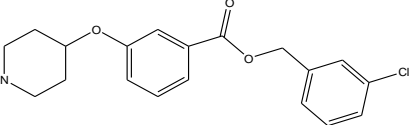
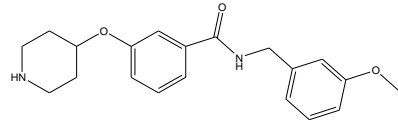
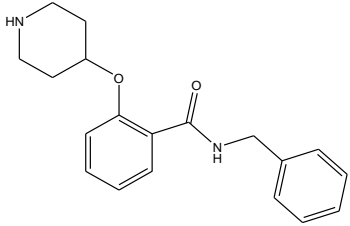
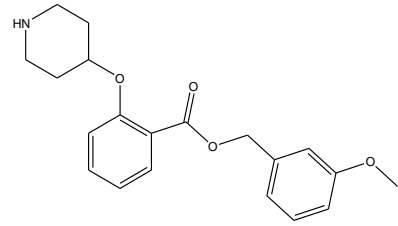
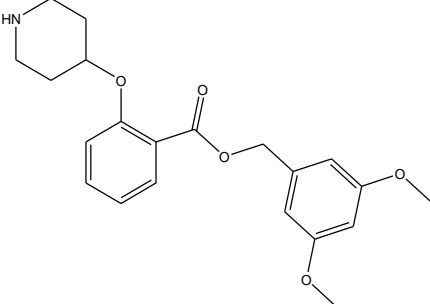
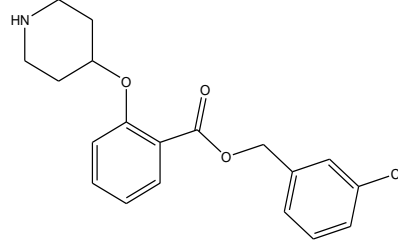
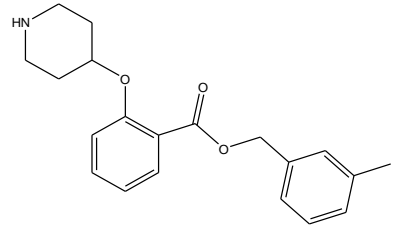
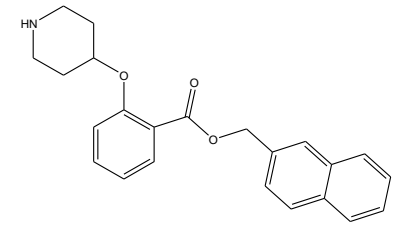
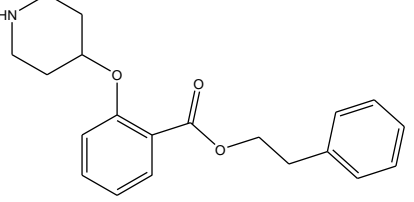
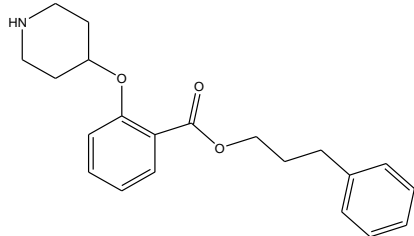
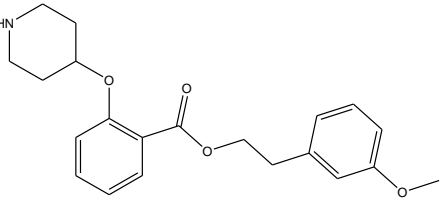
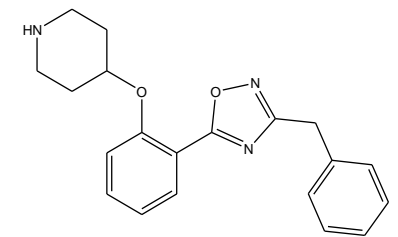
In this work, we used dimensional quantitative structure-activity relationship analysis of 83 NMT inhibitors seeking to propose new candidates for NMT inhibitors. Furthermore, an ADME evaluation was performed in order to find the most appropriate compound predicted.

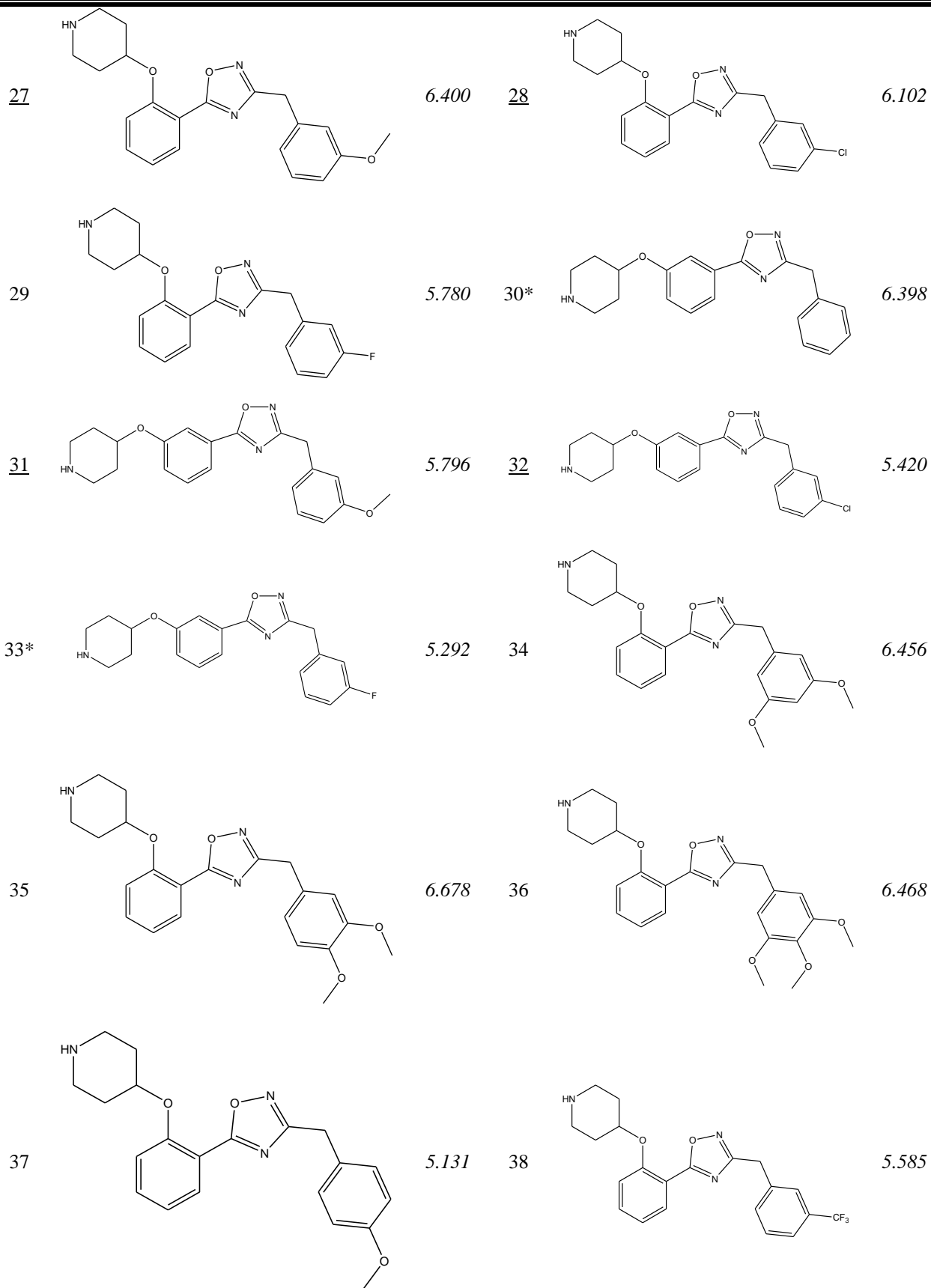
2 Methods

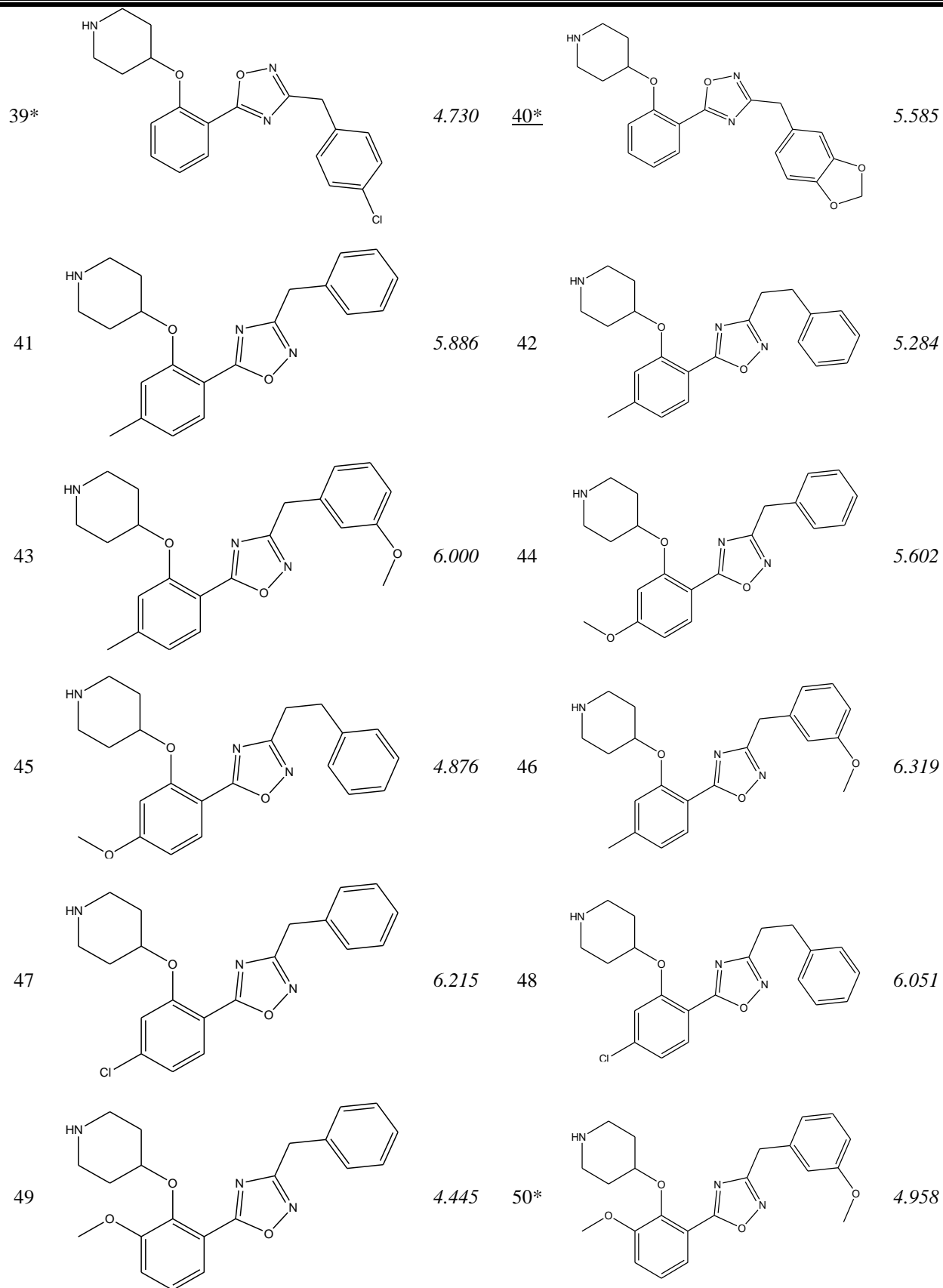
2.1 Biological Data

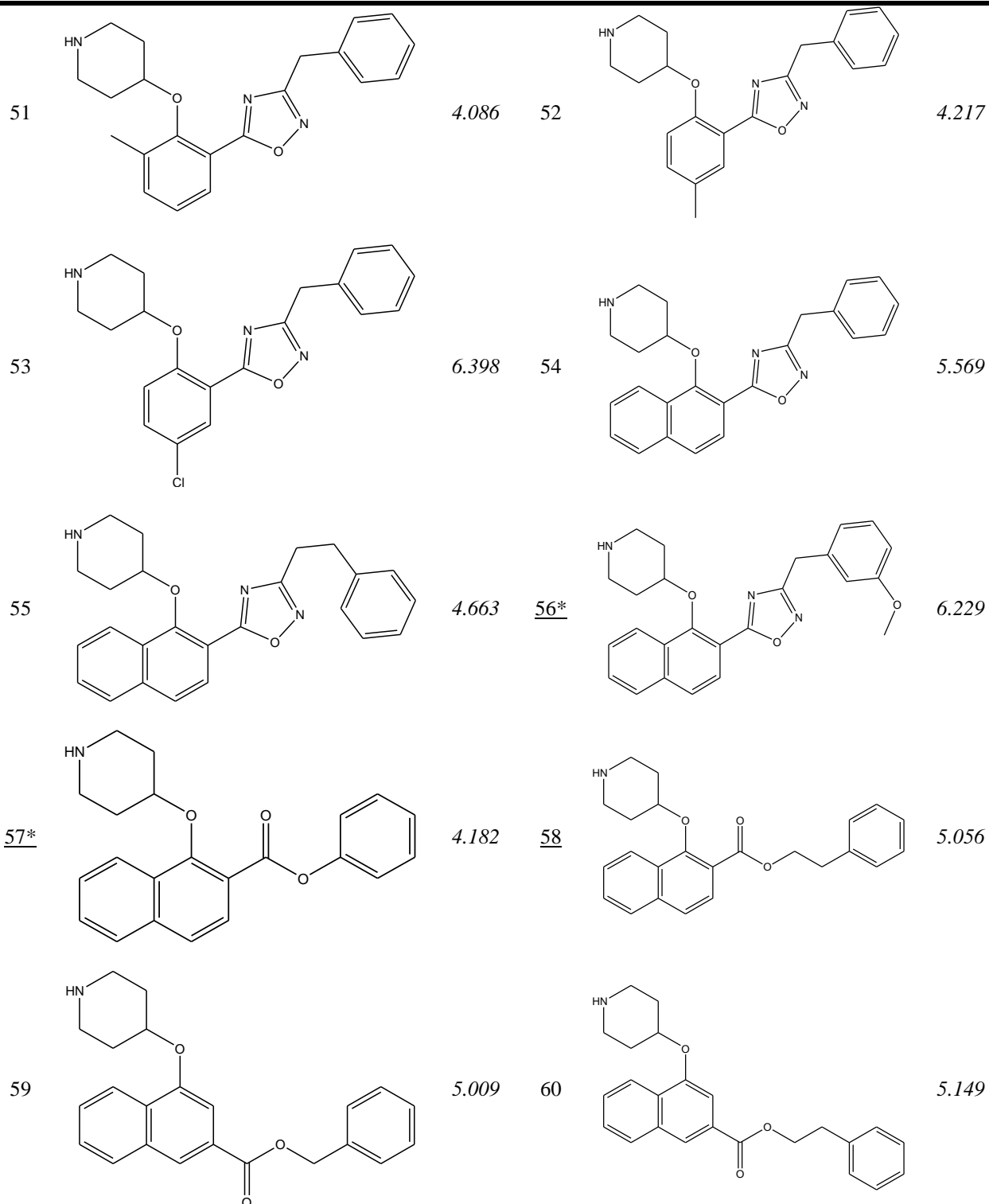
In order to build QSAR models, 83 *Plasmodium falciparum* inhibitors were retrieved from Leatherbarrow et al.[14]. 20 compounds (25%) were randomly selected to compose the test set (external validation). Two test groups were chosen. The first (Test set I) has the following molecules: **1; 3; 5; 6; 12; 16; 20; 30; 33; 39; 40; 50; 56; 57; 61; 65; 66; 69; 76 and 80**; Test set II has the following molecules: **3; 6; 9; 13; 20; 21; 27; 28; 31; 32; 40; 56; 57; 58; 64; 70; 73; 76; 78 and 82** (Fig. 1).

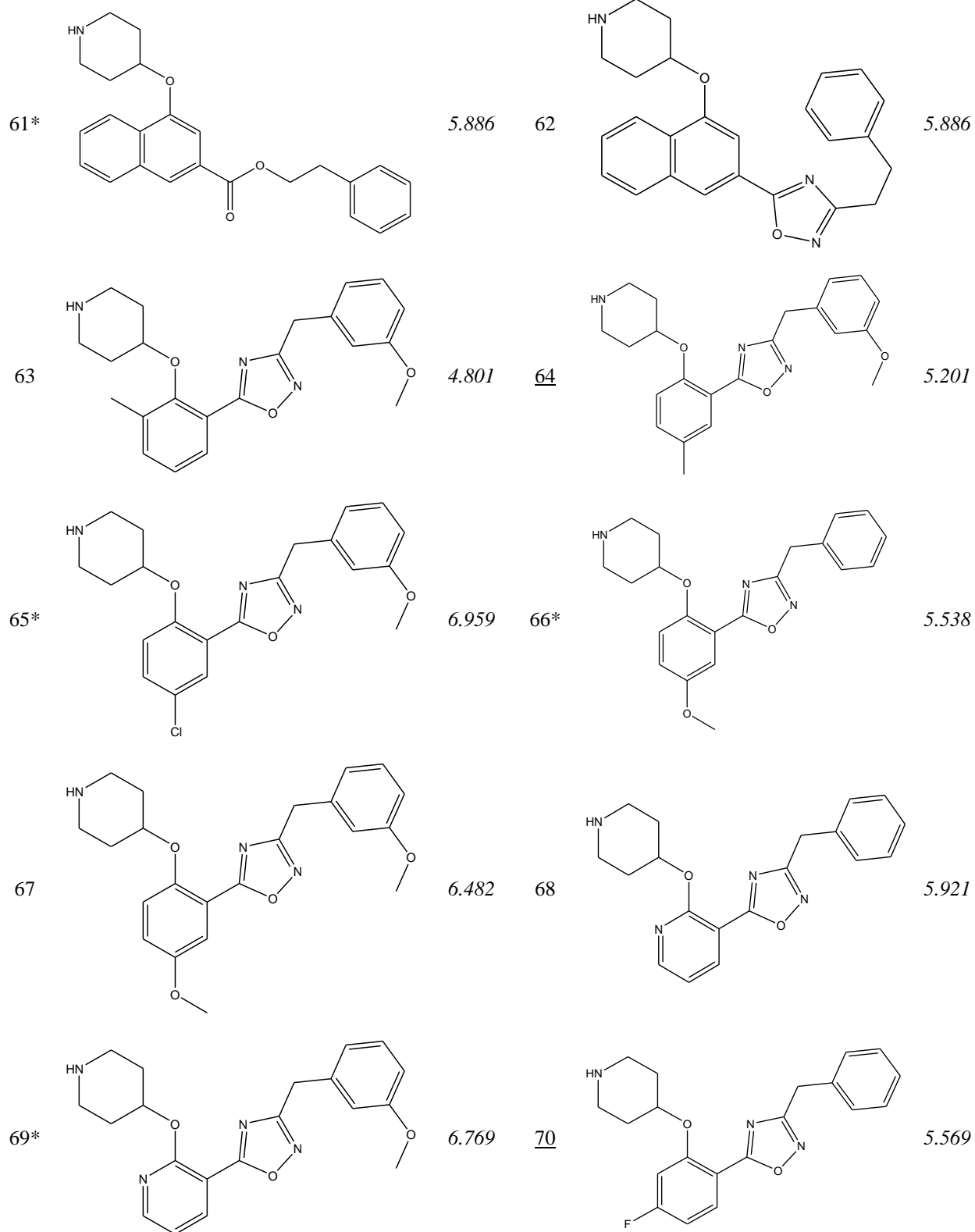
N°	Structure	pIC ₅₀	N°	Structure	pIC ₅₀
1*		6.155	2		4.000
3*		4.000	4		4.000
5*		4.000	6*		4.000
7		4.000	8		4.000
9		4.000	10		4.000
11		4.000	12*		5.721

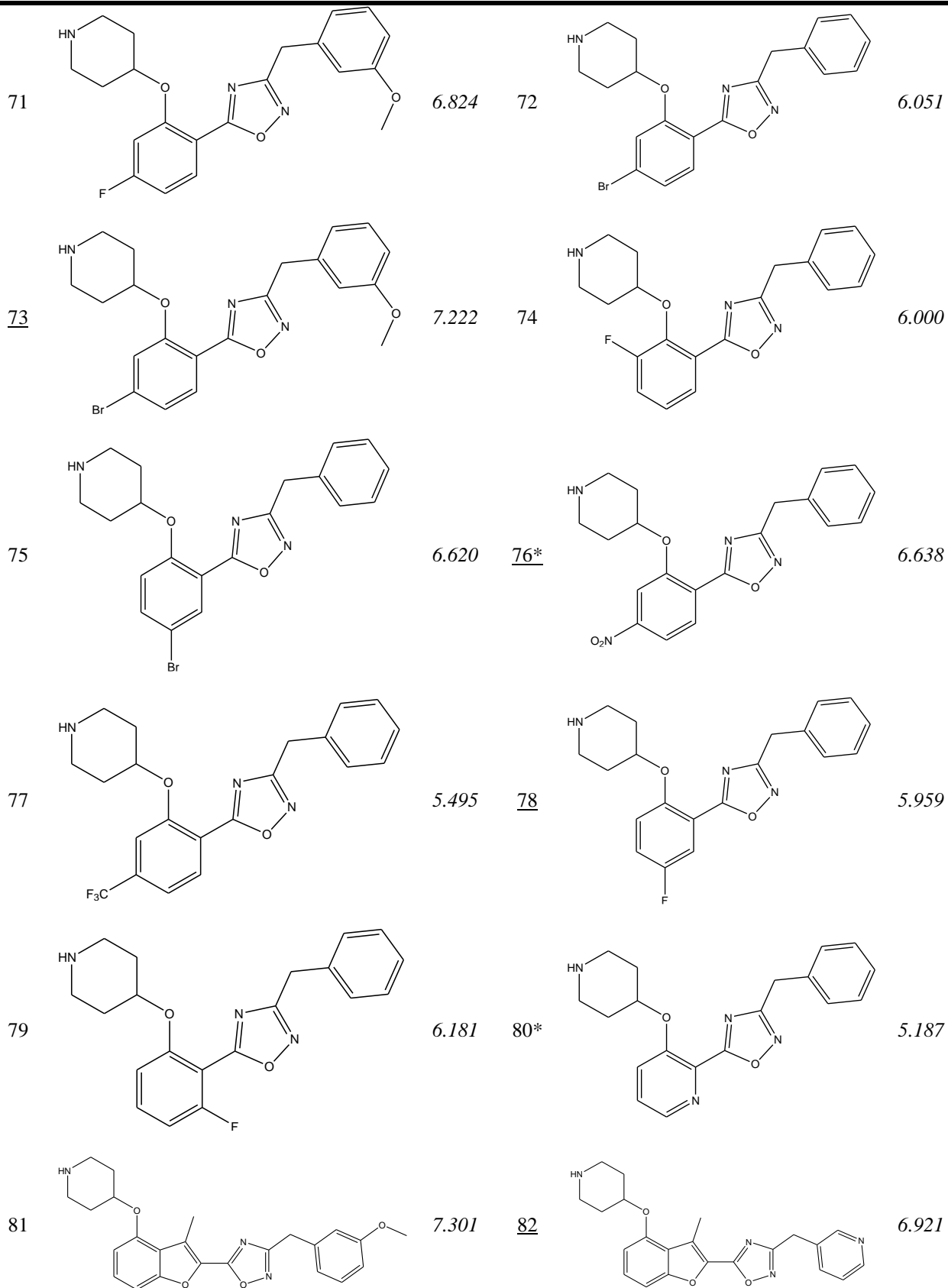
<u>13</u>		4.785	14		5.113
15		4.000	16*		4.000
17		4.000	18		4.745
19		5.215	<u>20*</u>		4.366
<u>21</u>		4.000	22		4.000
23		4.000	24		4.000
25		4.000	26		5.699











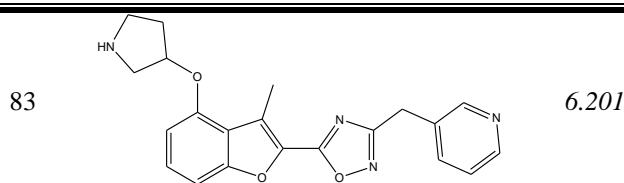


Fig 1. Chemical structures and pIC_{50} (M) values of 83 *Plasmodium falciparum* inhibitors. Test set I compound numbers are marked with an asterisk. Test set II compound numbers are underlined.

The biological activities of these compounds were reported as the concentration capable of inhibiting 50% of the enzyme activity (IC_{50}), measured using an adapted version of the sensitive fluorescence-based assay based on detection of CoA by 7-diethylamino-3-(4-maleimido-phenyl)-4-methylcoumarin [14]. Furthermore, in an effort to eliminate the potential noise that might have been introduced by the pooling of data sets from different sources, all pharmacological data are obtained from the same laboratory. The IC_{50} (μM) values were converted into molar units and then expressed in negative logarithmic units (pIC_{50}), and are depicted in Fig.1. The range of pIC_{50} values for the training and test set spans at least three orders of magnitude (4.00 to 7.30), and the biological activity values show a regular distribution over the whole range.

2.2 Molecular dynamic simulation (MDS)

The three-dimensional structure from 83 analogues (Fig. 1) were optimized in vacuum, without any restriction, and the partial atomic charges assigned using RM1 semiempirical Hamiltonian [15]. The MDS was carried out at 300 K, close to the temperature assays, with a simulation sampling time of 100 ps, and intervals of 0.001 ps. Thus, a total sample of 100,000 conformations of each compound was produced. The MDS calculations were carried out employing a distance-dependent dielectric function, $\epsilon_r = D^*r_{ij}$, which was set to 3^*r_{ij} in order to try to model the solvent effect.

2.3 Alignment definition

As the compounds are structural analogs, we will assume that all molecules bind to the receptor in a similar mode. In general, the alignments are chosen to span the common framework of the molecules in the training and test sets [16-18]. In this work, ten alignments were performed using atoms of the (common) benzene ring. Three-ordered atom trial alignments were selected: (1) a-b-c, (2) a-b-d (3) b-c-d, (4) c-d-f, (5) b-c-a, (6) b-a-c (7) a-c-b, (8)) c-b-a, (9) d-a-f and (10) a-b-e (Fig. 2). The order of the three ordered-atoms is important: the first atom

specified for a molecule might be expected to occupy a similar location in space as the first atom specified for the second molecule. The conformational ensemble profile (CEP) for each compound obtained after the MDS step was overlaid onto a cubic lattice with grid cell size of 1Å.

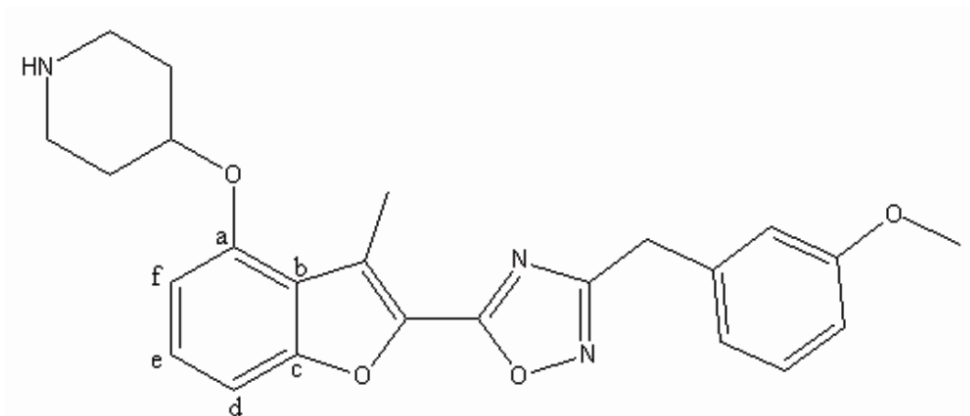


Fig. 2 Ordered atom letter codes (a, b, c, d, e and f) used in the 4D-QSAR analysis defines the three trial alignments. Compound **81** ($pIC_{50} = 7.301$) is used to define the atom letter code.

2.4 Interaction pharmacophore elements

According to the 4D-QSAR methodology, atoms of each compound are defined by seven types of interaction pharmacophore elements (IPEs). IPEs correspond to the interactions that may occur between ligand and the active site: i) any type (any); ii) nonpolar (np); iii) polar-positive charge density (p+); iv) polar-negative charge density (p-); v) hydrogen bond acceptor (hba); vi) hydrogen bond donor (hbd); and vii) aromatic systems (ar) (Hopfinger et al, 1997). The occupancy of the grid cells by each IPE type are recorded over the conformational assembly profile, and forms the set of grid cell occupancy descriptors (GCOD) to be utilized as the pool of trial descriptors in the model building and optimization process [19]. The idea underlying a 4D-QSAR analysis is that variations in biological responses were related to differences in the Boltzmann average spatial distribution of molecular shape with respect to the IPE. Thus, the normalized grid cell absolute occupancy, defined as the number of times that a cell was occupied by an atom type during the MDS, divided by the size of the CEP (1,000 conformations), was used to define the GCODs.

4D-QSAR model calculation. In order to exclude data noise from databases generated by the alignments, partial least-squares (PLS) regression analysis was performed as a data reduction fit between the observed dependent variable measures and the corresponding set of GCOD values. Additionally, PLS identifies the most highly weighted GCODs from data set of local grid cells [19].

200 GCODs with the highest weight from the data reduction were chosen to form the trial descriptor basis sets for model optimization by genetic function approximation (GFA) analysis [19]. Optimizations were initiated using 100 randomly generated models and 10,000-100,000 crossover operations. Mutation probability over the crossover optimization cycle was set at 10-30%. The smoothing factor, the variable that specifies the number of descriptors in the QSAR models, was varied between 1.0 and 3.0, in order to determine equations with no more than 12 terms. Each alignment was evaluated using the procedure described above.

The best models, resulting from the 4D-QSAR study were based on different criteria [20-24].

- i) Coefficient of determination (r^2): is a measure of how well the regression line represents the data.
- ii) Root mean square error of calibration (RMSE_C);
- iii) Leave-one-out cross-validation (LOOCV) correlation coefficient (q^2): estimating the performance of a predictive model.
- iv) Adjusted cross-validated squared correlation coefficient (q^2_{adj}):

$$q^2_{adj} = \frac{1-(1-q^2)}{\frac{n-1}{n-p}} \quad \text{Eq. 1}$$

- v) Root mean square error of cross validation (RMSE_{CV});
- vi) Least-square error of fit (LSE);
- vii) Friedman's lack-of-fit (LOF): estimates the quality of a model. This measure penalizes appropriately for the addition of terms to the equation (and consequent loss of degrees of freedom) in such a way to resist over-fitting:

$$LOF = \frac{LSE}{\left\{1 - \frac{(c+dp)}{m}\right\}^2} \quad \text{Eq.2}$$

where LSE is the least-squares error, c is the number of basis functions in the model, d is the smoothing factor, p is the total number of variables contained in all basis functions, and m is the number of samples (compounds) in the training set.

- viii) Correlation coefficient of external validation set (R^2_{pred}): reflects the degree of correlation between the observed ($Y_{Exp(test)}$) and predicted ($Y_{Pred(test)}$) activity data of the test set:

$$R^2_{Pred} = 1 - \frac{\sum_1^n (Y_{Exp(test)} - Y_{Pred(test)})^2}{\sum_1^n (Y_{Exp(test)} - \bar{Y}_{training})^2} \quad \text{Eq.3}$$

Where $\bar{Y}_{\text{training}}$ is average value for the dependent variable for the training set.

- ix) Root mean square error of prediction (RMSE_p),
- x) Modified r^2 ($r^2_{m(test)}$) equation determining the proximity between the observed and predicted values with the zero axis intersection:

$$r_m^2 = r^2 \left(1 - \left| \sqrt{r^2 - r_0^2} \right| \right) \quad \text{Eq.4}$$

- xi) Y-Randomization (R^2_r): consists of the exchange of the values of the independent variables randomly. Thus, the R^2_r value must be less than the correlation coefficient of the non-randomized models.
- xii) Root mean square error of Y-Randomization (RMSE_{C Y-rand});
- xiii) R^2_p : penalizes the model R^2 for the difference between the squared mean correlation coefficient (R^2_r) of randomized models and the square correlation coefficient (R^2) of the non-randomized model:

$$r_p^2 = r^2 * \sqrt{r^2 - r_r^2} \quad \text{Eq.5}$$

Conformational selection. In the 4D-QSAR method, the conformation of each compound can be postulated as the lowest-energy conformer state from the set sampled for each compound, which predicted the maximum activity using the optimum 4D-QSAR model [19].

3 Results

3.1 Alignments evaluation

The GA-PLS analysis using grid cells of 1.0 Å generated several models or equations. The statistical parameters of ten alignments studied for Test set I (Compounds **1, 3, 5, 6, 12, 16, 20, 30, 33, 39, 40, 50, 56, 57, 61, 65, 66, 69, 76** and **80**) and Test set II (Compounds **3, 6, 9, 13, 20, 21, 27, 28, 31, 32, 40, 56, 57, 58, 64, 70, 73, 76, 78** and **82**) are shown in Table 1 and Table 2, respectively. All tested alignments showed q^2 values higher than 0.5. This reveals that the model can be a useful tool for predicting affinities of new compounds based on these structures; r^2 greater than 0.7 indicates that the model is correlated and may be considered to represent the training set in the same manner [19]. Alignments 6B and 7B were eliminated from the analysis because it presented a low r^2 value (<0.7).

Evaluating the predictive ability in terms of R^2_{pred} (means of an external validation), the alignment 5A was eliminated ($R^2_{\text{pred}} < 0.5$). All R^2_{m} values were greater than 0.60, and values over 0.5 are acceptable. Analyzing the R^2_{p} values, alignments B1, B2, B4, B6 and B7 were excluded because this parameter value was less than 0.5 [25].

Alignment 3 from Test set II (B3) provides the best 4D-QSAR models as judged by the highest q^2_{adj} , in addition to presenting fewer descriptors. Among the alignments with only 7 descriptors, this still has the highest values of r^2 , q^2 , q^2_{adj} , R^2_{p} , and lowest value of R^2_{r} . According to these results, we selected Model B3 as the best alignment. We will only present the analysis of the best model derived from B3.

Table 1 Statistical parameters evaluated in the 4D-QSAR analysis for the ten performed alignments of the Test I.

Alignment	r^2	RMSE _C	q^2	q^2_{adj}	RMSE _{CV}	LSE	LOF	R^2_{Pred}	RMSE _P	R^2_m	R^2_r	RMSE _{C Y-rand}	R^2_p	GCODs
A1	0.746	0.481	0.684	0.607	0.549	0.243	0.374	0.532	0.650	0.710	0.312	0.492	0.820	8
A2	0.744	0.478	0.685	0.608	0.548	0.246	0.377	0.548	0.663	0.692	0.343	0.471	0.799	8
A3	0.761	0.469	0.687	0.609	0.546	0.229	0.398	0.508	0.702	0.735	0.182	0.579	0.994	8
A4	0.708	0.508	0.649	0.576	0.579	0.279	0.485	0.588	0.595	0.645	0.287	0.460	0.825	8
A5	0.736	0.511	0.664	0.589	0.566	0.253	0.439	0.477	0.698	0.766	0.245	0.516	0.895	8
A6	0.739	0.477	0.668	0.582	0.563	0.250	0.407	0.567	0.637	0.670	0.286	0.497	0.830	9
A7	0.722	0.503	0.658	0.584	0.571	0.266	0.462	0.555	0.656	0.683	0.291	0.474	0.831	8
A8	0.746	0.445	0.682	0.605	0.551	0.243	0.373	0.620	0.590	0.606	0.216	0.543	0.891	8
A9	0.734	0.491	0.664	0.578	0.566	0.255	0.363	0.547	0.684	0.693	0.250	0.511	0.861	9
A10	0.723	0.519	0.657	0.583	0.572	0.265	0.407	0.503	0.676	0.740	0.311	0.464	0.816	8

Table 2 Statistical parameters evaluated in the 4D-QSAR analysis for the ten performed alignments of the Test II.

Alignment	r^2	RMSE _C	q^2	q^2_{adj}	RMSE _{CV}	LSE	LOF	R^2_{Pred}	RMSE _P	R^2_{m}	R^2_{r}	RMSE _{C Y-rand}	R^2_{p}	GCODs
<i>B1</i>	0.728	0.504	0.683	0.617	0.544	0.261	0.377	0.728	0.532	0.688	0.301	0.817	0.476	7
<i>B2</i>	0.728	0.515	0.672	0.607	0.553	0.260	0.466	0.763	0.496	0.749	0.289	0.826	0.482	7
<i>B3</i>	0.757	0.472	0.702	0.634	0.527	0.233	0.418	0.746	0.515	0.716	0.110	1.055	0.609	7
<i>B4</i>	0.704	0.549	0.648	0.585	0.573	0.283	0.455	0.782	0.476	0.765	0.253	0.846	0.473	7
<i>B5</i>	0.725	0.500	0.675	0.601	0.550	0.264	0.473	0.706	0.553	0.692	0.198	0.930	0.526	7
<i>B6</i>	0.692	0.559	0.638	0.576	0.581	0.295	0.427	0.771	0.489	0.755	0.272	0.827	0.448	7
<i>B7</i>	0.690	0.556	0.643	0.581	0.577	0.297	0.533	0.751	0.509	0.735	0.289	0.825	0.437	7
<i>B8</i>	0.730	0.514	0.676	0.600	0.550	0.259	0.397	0.770	0.489	0.750	0.209	0.877	0.527	8
<i>B9</i>	0.723	0.528	0.670	0.605	0.555	0.265	0.426	0.786	0.472	0.773	0.229	0.873	0.508	7
<i>B10</i>	0.744	0.501	0.685	0.619	0.542	0.246	0.394	0.779	0.480	0.744	0.289	0.818	0.502	7

The statistical measures, including the values of r^2 , q^2 , q^2_{adj} , LSE, LOF, RMSE_C, RMSE_{CV}, RMSE_P, RMSE_C, R^2_{Y-Rand} , R^2_{pred} , R^2_m , R^2_p and R^2_r , are presented below. Each GCOD is labeled as “x, y, z, IPE” which represent the cartesian coordinate positions of the selected grid cell (x, y, z) and the respective atom type (IPE).

Equation 6 (Model B3)

$$pIC_{50} = 3.997 + 1.191(-1, -4, -3, any) - 8.269(-1, -4, -4, np) + 2.100(0, -1, 0, any) \\ + 0.846(0, 6, 2, any) + 2.345(0, -5, -1, any) + 1.692(0, 3, -3, any) \\ + 4.942(0, -3, -1, hba)$$

$n = 63$, $GCODs = 7$, $r^2 = 0.757$, $q^2 = 0.702$, $q^2_{adj} = 0.634$, $LSE = 0.233$, $LOF = 0.418$, $RMSE_C = 0.472$, $RMSE_{CV} = 0.527$, $RMSE_P = 0.515$, $RMSE_{C_{y-rand}} = 1.055$, $R^2_{pred} = 0.746$, $R^2_m = 0.716$, $R^2_p = 0.609$, $R^2_r = 0.110$.

Another different variant of R^2_m metrics was calculated from Model B3 to assess the predictive ability of test set, $\Delta R^2_{m[23]}$. The value of ΔR^2_m found was 0.133. It has been suggested that to be considered a predictive model, this value should be less than 0.2. Model B3 generated seven descriptors (Equation 6), where GCODs (-1,-4,-3,any), (0,-1,0,any), (0,6,2,any), (0,-5,-1,any), (0,3,-3,any), (0,-3,-1,hba) present positive coefficients (Eq. 6) and correspond to favorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Therefore, substituents in these positions increase the effectiveness of the compounds. The GCOD (-1,-4,-4, np) has negative coefficient and correspond to unfavorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Therefore, the occupation of GCOD (-1,-4,-4, np) decreases the compound potency.

4 Discussion

GCODs are related to the coordinates of IPE mapped in a common grid. A graphic representation of the descriptors of Model B3 is shown in Fig. 3 using Compound **81** as a reference. Light and dark spheres represent GCODs with positive and negative coefficients, respectively, in accordance with Eq.6 (Model B3).

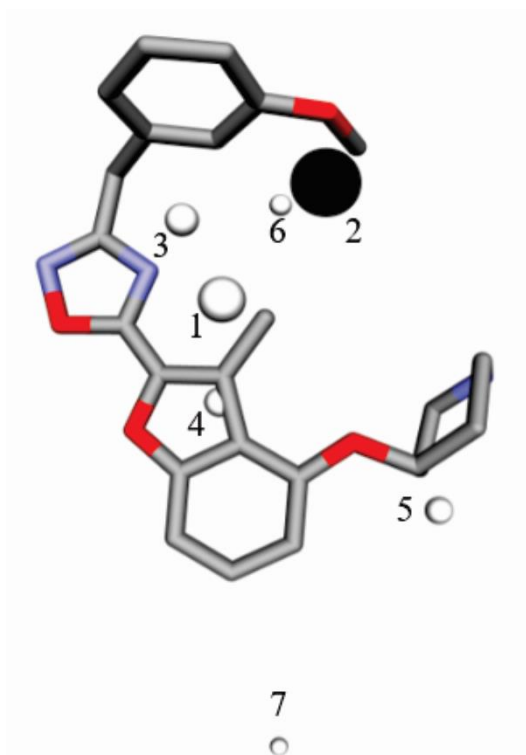


Fig. 3 Graphic representation of Compound **81** according to the 4D-QSAR Model B3. GCODs occupancy represented by white spheres contribute to increasing the potency of compounds, and black spheres to decrease the potency of the compounds. The GCODs described are: (1) (0,-3,-1, hba), (2) (-1,-4,-4, np), (3) (0,-5,-1, any) (4) (0,-1,0, any), (5) (0,3,-3, any) (6) (-1,-4,-3, any), (7) (0,6,2, any).

GCOD-1 (0,-3,-1,hba) (Fig. 4) is the descriptor that most contributes to the increased effectiveness of compounds and presents a coefficient of 4.942 (Eq.6). This grid cell represents an acceptor hydrogen bond atom type (IPE) and shows high frequency of occupation for Compounds **42**, **48**, **65**, **68** and **69**. It is located close to the nitrogen atom of the oxadiazole ring and indicates an aminoacid donor hydrogen bond in N-myristoyltransferase.

The oxadiazole ring in the ortho position allows the nitrogen atom to occupy this grid cell, exemplified by Compound **42** (Fig. 4). However, the most active molecules of the training set, **81** and **83**, do not have this descriptor that most contributes to the increase in the potency of the compounds, once in this position there is a methyl group and the oxadiazole group was displaced. In fact, the oxadiazole ring in these compounds do not occupy this grid cell.

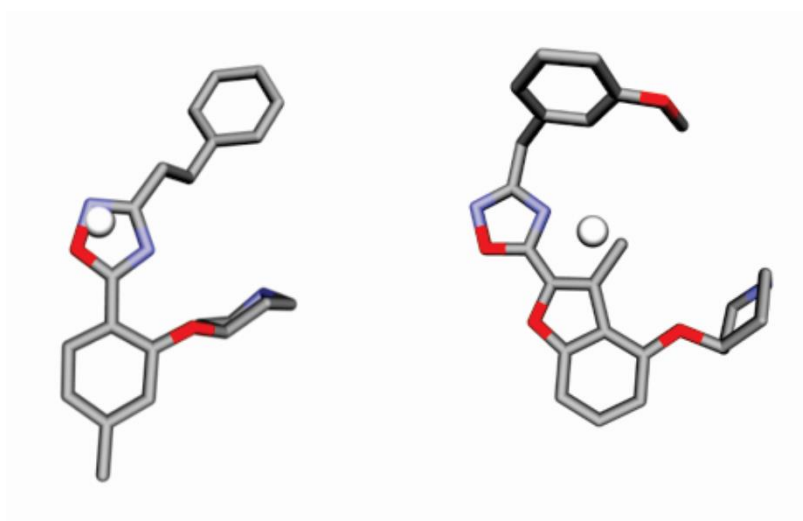


Fig. 4 A) Representation of Compound **42** and GCOD-1 (0,-3,-1,hba) (white sphere). **B)** Compound **81**.

GCOD-2 (-1,-4,-4,np) (Fig. 5) contributes to decrease compound potency and presents a coefficient of -8.269. This grid cell corresponds to a nonpolar IPE and shows high occupation frequency for Compounds **42**, **48** and **55**. These molecules present non-polar groups in this local, such as ethyl and methyl groups. Thus, the occupation of this cell is drastically reduced when this position is non-polar substituted, that negatively influences activity of these compounds. We can see that Compound **42** has occupation in the descriptor that discourages the activity, however, it also presents occupancy for GCOD-1, which favors the activity. As expected, this compound exhibits an intermediate power.

Instead, molecules **81** and **83** (more active) have oxygen and nitrogen atoms, respectively, in this grid cell and, thus, have low coefficients for GCOD (-1, -4, -4, np). This suggests that the occupation of this cell by acceptor hydrogen bond should be favorable in contrast to nonpolar atoms that are unable to perform hydrogen bond interactions (Fig. 5).

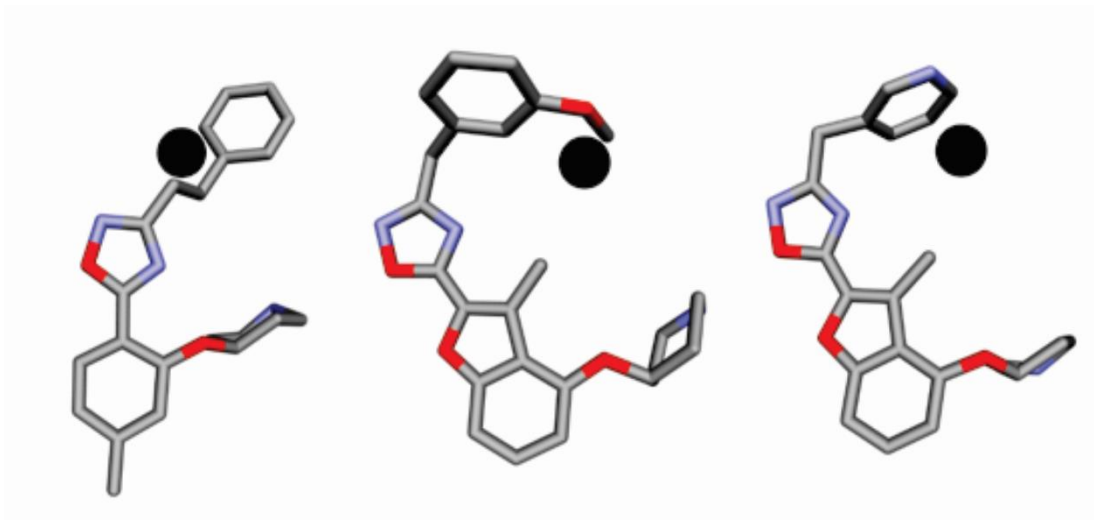


Fig. 5 Representation of GCOD-2 (-1,-4,-4,np) (black sphere) and **A**) Compound **42** **B**) Compound **81** **C**) Compound **83**.

GCOD-3 (0,-5,-1,any) is present as a non-specific IPE. The atom type “any” is used when more than one specific atom [IPE] type across the training set is found to satisfy the interaction being captured by a particular GCOD [26]. GCOD (0,-5,-1, any) has a positive coefficient of 2,345 which increases the potency of the compounds (Fig. 6). This grid cell is close to methyl and ethyl groups, toward the left side, and shows greater occupation frequency for Compounds **26**, **30** and **33**. On the other hand, Compounds **42** and **45** showed no occupancy for this descriptor, because during the molecular dynamics simulation, they assumed a different conformation, facing right. On the other hand, GCOD (0,-5,-1,any) also provides occupation frequency in the benzene ring of Molecules **59**, **60**, **61** and **62**.

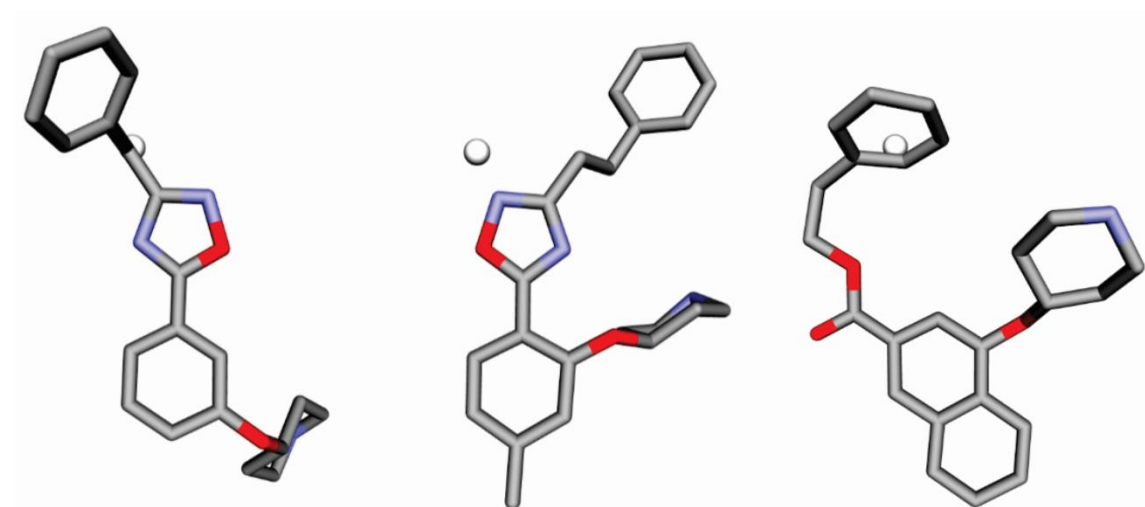


Fig. 6 Representation of GCOD-3 (0,-5,-1,any) (white sphere) and **A**) Compound **30** **B**) Compound **42** **C**) Compound **60**.

GCOD-4 (0,-1,0,any) have a positive coefficient and thus also greatly influence the increase in inhibitor potency (Fig. 7). It is located near the methyl group in benzofuran, 2,3-dihydro-3-methyl and represents a non-specific IPE. It shows the highest occupation frequency for the most active compounds in this series, Compounds **1**, **81** and **83**. The GCOD (0,-3,-1,hba) reflect the importance of occupation of this receptor region for the effectiveness of the inhibitors.

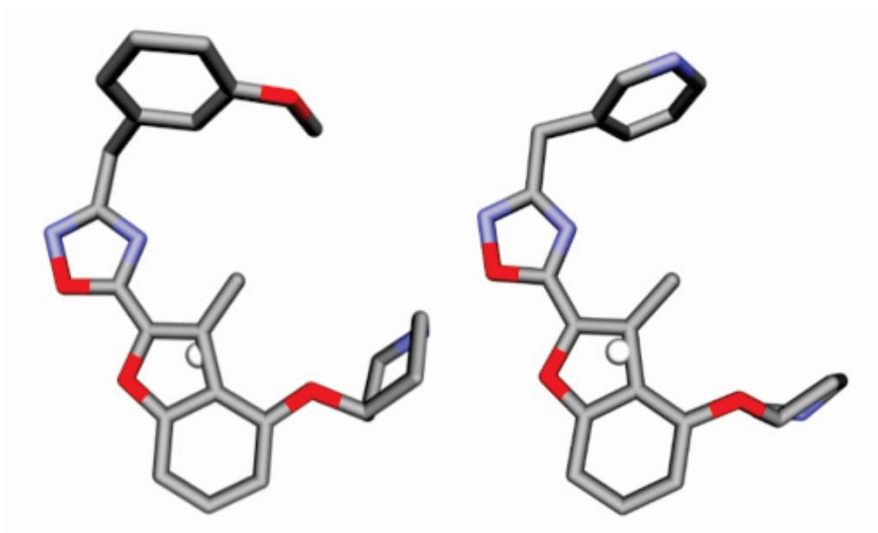


Fig. 7 Representation of GCOD-4 (0,-1,0,any) (white sphere) and **A) Compound 81 B) Compound 83**.

GCOD-5 (0,3,-3,any) have a positive coefficient, and so, improves the effectiveness of inhibitors (Fig. 8). The GCOD (0,3,-3,any) is situated near the piperidine ring and has occupancy for a large majority of molecules, such as Compounds **65** and **79**. The change of position of the piperidine ring in Compounds **15**, **16** and **30** is not favorable, so these do not have this GCOD.

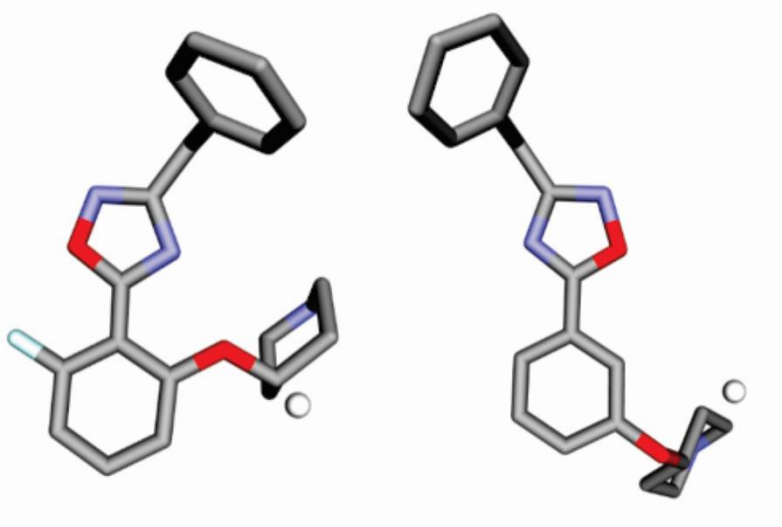


Fig. 8 Representation of GCOD-5 (0,3,-3,any) (white sphere) and **A) Compound 79 B) Compound 30**.

GCOD-6 (-1,-4,-3,any) represents a non-specific IPE and also has a positive coefficient, indicating an increase in potency of compounds which have high occupation frequency for this descriptor (Fig. 9). The Molecules **14**, **19** and **22** have a high occupation frequency for this GCOD, located near the benzene group. This grid cell suggests a hydrophobic region in the receptor close to benzene ring, which should be making a π - π staking interaction between the aromatic ring of an inhibitor and one aromatic amino acid residue.

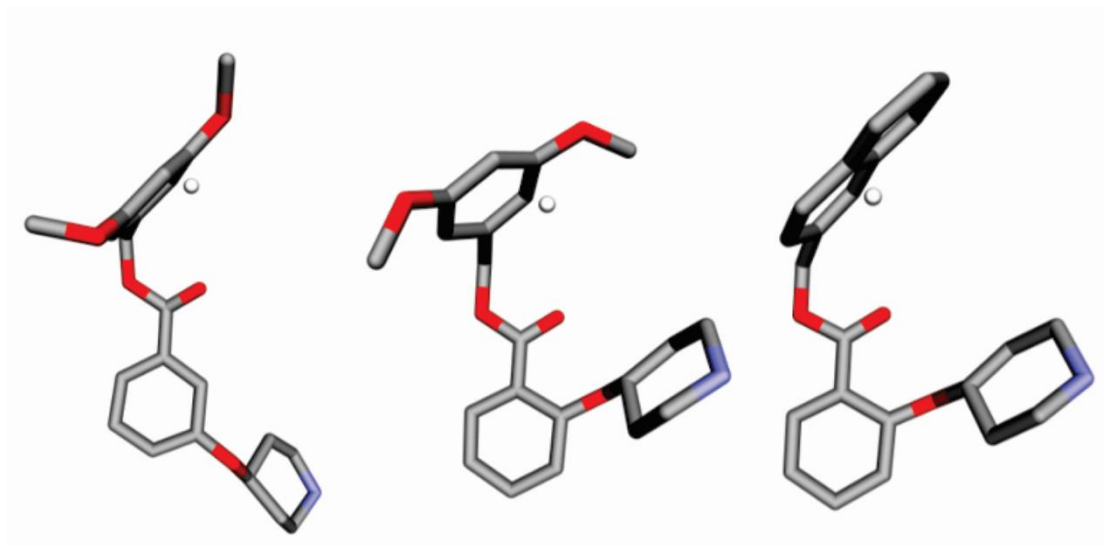


Fig. 9 Representation of GCOD-6 (-1,-4,-3, any) (white sphere) and **A)** Compound **14** **B)** Compound **19** **C)** Compound **22**.

Lastly, GCOD-7 (0,6,2,any) (Fig. 10) has a positive coefficient and shows a non-specific class. Molecules **59**, **60**, **61** and **62** and **63** have a high occupation frequency for this descriptor. The presence of naphthalene group in this grid cell is favorable to the activity. In fact, it shows that the aromatic substituents in this position should be preferred.

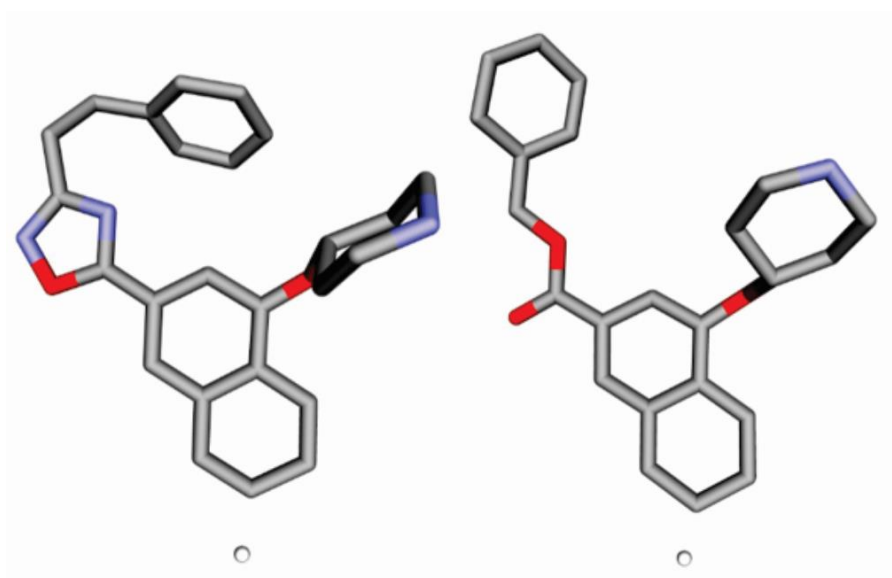
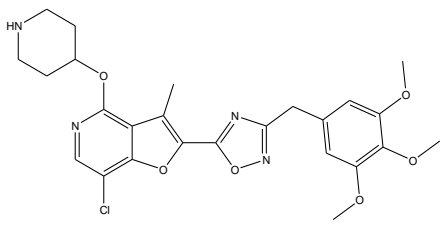
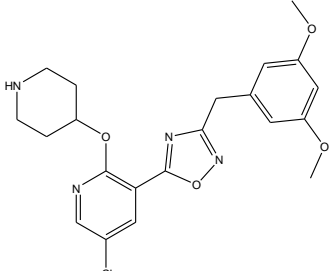
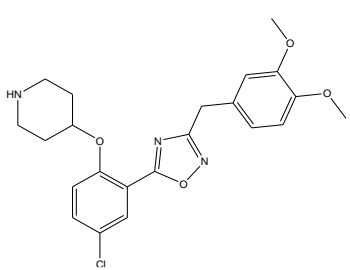
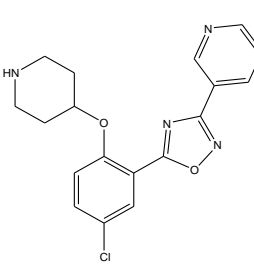
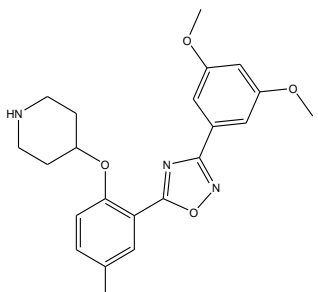


Fig. 10 Representation of GCOD-7 (0,6,2,any) (white sphere) and **A) Compound 62 B) Compound 59.**

The analysis of the 4D-QSAR result allowed to observe the main characteristics that a potent NMT inhibitor should present. Thus, the descriptor information obtained by Model B3 was used. Modifications in the structure of Compounds **60**, **65** and **81** were suggested and Compounds **A**, **B**, **C**, **D** and **E** were proposed.

The proposed compounds exhibited predicted pIC_{50} higher than **81** (the best experimental compound), whose predicted pIC_{50} was 6.94. The structure of the five compounds and their predicted pIC_{50} values are shown in Table 3, where it is possible to verify that compound **D** presented the best predicted potency. The ADME (absorption, distribution, metabolism and excretion) of a drug is an important property that can determine the utilization of molecules proposed in the therapeutic usage. For the evaluation of pharmacokinetic parameters, Molecules **A**, **B**, **C**, **D** and **E** we used the Lipinski's rule of five, where molecular properties are closely related to the oral bioavailability of a drug [27], wherein compounds should not violate more than one rule. In this rule, the compounds should present $\log P \leq 5$, molecular weight ≤ 500 , number of hydrogen bond acceptors (n_{ON}) ≤ 10 , number of hydrogen bond donors (n_{OHNH}) ≤ 5 , and number of rotatable bonds (n_{rotb}) ≤ 10 . The proposed molecules have been designed in the Molinspiration Online Property Calculation Software Toolkit (<http://www.molinspiration.com/>) to evaluate the criteria discussed above.

Table 3. Structures of the proposed compounds and the predicted pIC₅₀ values based on Model B3.

N°	Structure	pIC ₅₀	N°	Structure	pIC ₅₀
A		7.014	B		7.171
C		7.622	D		8.161
E		7.894			

The Molinspiration Online Property Calculation Software Toolkit (<http://www.molinspiration.com/>) was used to analyze drug likeness (Lipinski's Rule of Five) and results are shown in Table 4. According to the data in Table 4, no compound violated the Lipinski's Rule of Five.

Table 4. Calculated parameters of the Lipinski rule of five for the proposed molecules.

Molecule	miLogP	MW	n _{ON}	n _{OHNH}	n	n _{violations}
A	3.13	514.97	10	1	8	1
B	2.74	430.89	8	1	7	0
C	3.13	415.88	7	1	6	0
D	2.41	356.81	6	1	4	0
E	3.52	415.88	7	1	6	0

5 Conclusions

In summary, 4D-QSAR models for NMT inhibitors were built and evaluated. Two test groups were evaluated for the ten tested alignments. The best model was obtained from Alignment B3, and generated an equation with seven descriptors, six of which have positive coefficients and one a negative coefficient. Model B3 showed a satisfactory statistical quality and predictive abilities as shown by $r^2 = 0.757$, $q^2 = 0.702$, $q^2_{\text{adjusted}} = 0.634$, $R^2_{\text{pred}} = 0.746$, $R^2_{\text{m}} = 0.716$, $R^2_{\text{p}} = 0.609$. Furthermore, it showed low values of $R^2_{\text{r}} = 0.110$, and $\Delta R^2_{\text{m}(\text{test})} = 0.133$. 4D-QSAR analysis indicated an important role of acceptor hydrogen bonding groups and aromatic groups, allowing to propose five structures. These have proved more active than Compound **81**, in addition to being assessed by the Lipinski's Rule. Accordingly, these molecules may be considered promising prototypes against malaria.

Acknowledgements. We thank the Brazilian agencies "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq, Brazil), "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) and "Fundação de Amparo à Pesquisa do Estado de Minas Gerais" (FAPEMIG, Brazil) for their support. We thank A. J. Hopfinger who kindly supplied the 4D-QSAR program for academic use.

References

1. Njoroge M, Njuguna NM, Mutai P, Ongarora DSB, Smith PW, Chibale K (2014) Recent Approaches to Chemical Discovery and Development Against Malaria and the Neglected Tropical Diseases Human African Trypanosomiasis and Schistosomiasis. *Chemical Reviews* 114:11138-11163.
2. Liu W, Li Y, Learn GH, Rudicell RS et al (2010) Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467:420-U467.
3. World malaria report (2015) World Health Organization, Geneva <http://www.who.int/malaria/publications/world-malaria-report-2015/report/en/>. Accessed 25 July 2016.
4. Goldston AM, Sharma AI, Paul KS, Engman DM (2014) Acylation in trypanosomatids: an essential process and potential drug target. *Trends in Parasitology* 30:350-360.

5. Rudnick DA, McWherter CA, Rocque WJ, Lennon PJ, Getman DP, Gordon JI (1991) kinetic and structural evidence for a sequential ordered bi bi mechanism of catalysis by *Saccharomyces cerevisiae* myristoyl-coa - protein n-myristoyltransferase. *Journal of Biological Chemistry* 266:9732-9739.
6. Tate EW, Bell AS, Rackham MD, Wright AH (2014) N- Myristoyltransferase as a potential drug target in malaria and leishmaniasis. *Parasitology* 141:37-49.
7. Gunaratne RS, Sajid M, Ling IT, Tripathi R, Pachebat JA, Holder AA (2000) Characterization of N-myristoyltransferase from *Plasmodium falciparum*. *Biochemical Journal* 348:459-463.
8. Krungkrai SR, Krungkrai J (2016) Insights into the pyrimidine biosynthetic pathway of human malaria parasite *Plasmodium falciparum* as chemotherapeutic target. *Asian Pacific Journal of Tropical Medicine* 9:525-534.
9. Ridley RG (2002) Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* 415:686-693.
10. Fidock DA, Nomura T, Talley AK, Cooper RA, Dzekunov SM, Ferdig MT, Ursos LMB, Sidhu ABS, Naude B, Deitsch KW, et al (2000) Mutations in the P-falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Molecular Cell* 6:861-871.
11. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, Lwin KM, Ariey F, Hanpithakpong W, Lee SJ, et al (2009) Artemisinin Resistance in *Plasmodium falciparum* Malaria. *New England Journal of Medicine* 361:455-467.
12. Okombo J, Chibale K (2016) Antiplasmodial drug targets: a patent review (2000-2013). *Expert Opinion on Therapeutic Patents* 26:107-130.
13. Wright MH, Heal WP, Mann DJ, Tate EW (2010) Protein myristoylation in health and disease. *Journal of Chemical Biology*, 3:19-35.
14. Leatherbarrow R, et al (2013) Novel Compounds and their use in therapy. Imperial Innovations Limited. London Patent WO 2013/083991 13 June 2013.
15. Rocha GB, Freire RO, Simas AM, Stewart JJP (2006) RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *Journal of Computational Chemistry* 27:1101-1111.
16. Oliveira AA, Ramalho TC, da Cunha EFF (2009) QSAR Study of Androstenedione Analogs as Aromatase Inhibitors. *Letters in Drug Design & Discovery* 6:554-562.
17. da Cunha EFF, Albuquerque MG, Antunes OAC, de Alencastro RB (2005) 4D-QSAR models of HOE/BAY-793 analogues as HIV-1 protease inhibitors. *Qsar & Combinatorial Science* 24:240-253.
18. Caldas GB, Ramalho TC, da Cunha EFF (2014) Application of 4D-QSAR studies to a series of benzothiophene analogs. *Journal of Molecular Modeling* 20.
19. Hopfinger AJ, Wang S, Tokarski JS, Jin BQ, Albuquerque M, Madhav PJ, Duraiswami C (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *Journal of the American Chemical Society* 119:10509-10524.
20. Kiralj R, Ferreira MMC (2009) Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *Journal of the Brazilian Chemical Society* 20:770-787.
21. Kar S, Roy K (2011) Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. *Indian Journal of Biochemistry & Biophysics* 48:111-122.
22. Roy K, Paul S (2009) Exploring 2D and 3D QSARs of 2,4-Diphenyl-1,3-oxazolines for Ovicidal Activity Against *Tetranychus urticae*. *Qsar & Combinatorial Science* 28:406-425.
23. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) Some case studies on application of "r(m)(2)" metrics for judging quality of quantitative structure-activity relationship predictions: Emphasis on scaling of response data. *Journal of Computational Chemistry* 34:1071-1082.
24. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal R K (2011) Validation of QSAR Models - Strategies and Importance *International Journal of Drug Design & Discovery* 3:511-519.

25. Roy PP, Paul S, Mitra I, Roy K (2009) On Two Novel Parameters for Validation of Predictive QSAR Models. *Molecules* 14:1660-1701.
26. Pan DH, Liu JZ, Senese C, Hopfinger AJ, Tseng Y (2004) Characterization of a ligand-receptor binding event using receptor-dependent four-dimensional quantitative structure-activity relationship analysis. *Journal of Medicinal Chemistry* 47:3075-3088.
27. Patel HM, Sing B, Bhardwaj V, Palkar M, Shaikh MS, Rane R, Alwan WS, Gadad AK, Noolvi MN, Karpoormath R (2015) Design, synthesis and evaluation of small molecule imidazo 2,1-b 1,3,4 thiadiazoles as inhibitors of transforming growth factor-beta type-I receptor kinase (ALK5). *European Journal of Medicinal Chemistry* 93:599-613.

Supporting Information Available

Table S1 Descriptors generated by Model B3.

	-1,-4,-3,any	-1,-4,-4,np	0,-1,0,any	0,6,2,any	0,-5,-1,any	0,3,-3,any	0,-3,-1,hba
1	0.000	0.000	0.982	0.000	0.000	0.127	0.000
2	0.000	0.000	0.000	0.000	0.000	0.056	0.000
4	0.000	0.000	0.000	0.000	0.011	0.000	0.000
5	0.000	0.000	0.002	0.000	0.000	0.238	0.000
7	0.467	0.010	0.001	0.030	0.001	0.001	0.000
8	0.006	0.000	0.002	0.000	0.000	0.035	0.000
10	0.393	0.000	0.000	0.000	0.000	0.006	0.000
11	0.006	0.000	0.000	0.000	0.000	0.010	0.000
12	0.457	0.000	0.001	0.000	0.000	0.000	0.000
14	0.519	0.003	0.000	0.000	0.002	0.004	0.000
15	0.002	0.000	0.000	0.000	0.002	0.000	0.000
16	0.002	0.000	0.000	0.000	0.003	0.000	0.000
17	0.003	0.000	0.000	0.000	0.039	0.002	0.000
18	0.386	0.017	0.000	0.000	0.007	0.004	0.000
19	0.479	0.000	0.000	0.000	0.002	0.004	0.000
22	0.405	0.004	0.001	0.000	0.005	0.004	0.000
23	0.124	0.014	0.000	0.000	0.041	0.040	0.000
24	0.004	0.000	0.000	0.000	0.004	0.005	0.000
25	0.145	0.013	0.000	0.000	0.025	0.004	0.000
26	0.000	0.000	0.000	0.000	0.771	0.041	0.000
29	0.000	0.000	0.001	0.000	0.741	0.003	0.000
30	0.000	0.000	0.000	0.000	0.845	0.000	0.000
33	0.000	0.000	0.000	0.000	0.837	0.005	0.000
34	0.000	0.000	0.000	0.000	0.686	0.008	0.000
35	0.000	0.000	0.000	0.000	0.744	0.100	0.000

36	0.000	0.000	0.000	0.000	0.717	0.060	0.000
37	0.000	0.000	0.000	0.000	0.678	0.000	0.000
38	0.000	0.000	0.000	0.000	0.740	0.141	0.000
39	0.000	0.000	0.001	0.000	0.737	0.004	0.000
41	0.046	0.219	0.001	0.000	0.000	0.005	0.652
42	0.065	0.307	0.002	0.000	0.000	0.060	0.735
43	0.075	0.209	0.004	0.000	0.000	0.001	0.618
44	0.044	0.186	0.001	0.041	0.000	0.003	0.653
45	0.050	0.250	0.003	0.077	0.000	0.012	0.661
46	0.082	0.198	0.002	0.088	0.000	0.012	0.671
47	0.018	0.223	0.000	0.000	0.000	0.013	0.678
48	0.049	0.268	0.002	0.000	0.000	0.004	0.711
49	0.067	0.231	0.005	0.000	0.000	0.003	0.573
50	0.094	0.246	0.005	0.000	0.000	0.018	0.625
51	0.040	0.266	0.012	0.000	0.000	0.001	0.494
52	0.021	0.238	0.002	0.000	0.000	0.006	0.664
53	0.046	0.191	0.001	0.000	0.000	0.005	0.685
54	0.017	0.126	0.007	0.774	0.000	0.000	0.545
55	0.015	0.375	0.002	0.798	0.000	0.003	0.558
59	0.000	0.000	0.150	0.814	0.020	0.000	0.000
60	0.216	0.027	0.186	0.786	0.049	0.000	0.000
61	0.000	0.000	0.300	0.848	0.031	0.000	0.000
62	0.077	0.000	0.302	0.857	0.129	0.000	0.000
63	0.396	0.138	0.021	0.000	0.000	0.033	0.361
65	0.105	0.200	0.003	0.000	0.000	0.565	0.602
66	0.103	0.229	0.002	0.000	0.000	0.496	0.615
67	0.117	0.207	0.003	0.000	0.000	0.538	0.626
68	0.068	0.279	0.001	0.000	0.000	0.277	0.746

69	0.055	0.255	0.000	0.000	0.000	0.282	0.788
71	0.134	0.244	0.002	0.000	0.000	0.598	0.582
72	0.136	0.237	0.004	0.000	0.000	0.549	0.606
74	0.146	0.210	0.006	0.000	0.000	0.433	0.503
75	0.144	0.198	0.004	0.000	0.000	0.566	0.614
77	0.109	0.254	0.005	0.000	0.000	0.545	0.608
79	0.123	0.235	0.000	0.000	0.000	0.604	0.650
80	0.137	0.275	0.004	0.000	0.000	0.082	0.644
81	0.000	0.010	0.961	0.000	0.000	0.597	0.000
83	0.000	0.000	0.986	0.000	0.000	0.297	0.000

Table S2 Experimental and predicted pIC₅₀ values and residuals values (pIC₅₀predicted - pIC₅₀experimental) for compounds of Model B3.

Compound	pIC₅₀ Experimental	pIC₅₀ Predicted	Residuals
1	6.155	6.275	0.120
2	4.000	4.092	0.092
3	4.000	4.128	0.128
4	4.000	4.023	0.023
5	4.000	4.404	0.404
6	4.000	4.691	0.691
7	4.000	4.502	0.502
8	4.000	4.068	0.068
9	4.000	4.039	0.039
10	4.000	4.475	0.475
11	4.000	4.021	0.021
12	5.721	4.543	-1.178
13	4.785	4.012	-0.773
14	5.113	4.602	-0.511
15	4.000	4.004	0.004
16	4.000	4.007	0.007
17	4.000	4.096	0.096
18	4.745	4.339	-0.406
19	5.215	4.579	-0.636
20	4.366	4.449	0.083
21	4.000	4.102	0.102
22	4.000	4.467	0.467
23	4.000	4.193	0.193
24	4.000	4.020	0.020
25	4.000	4.128	0.128
26	5.699	5.874	0.175
27	6.400	5.899	-0.502
28	6.102	5.836	-0.267
29	5.780	5.742	-0.038
30	6.398	5.979	-0.420
31	5.796	5.828	0.032
32	5.420	5.961	0.541
33	5.292	5.968	0.676
34	6.456	5.619	-0.837
35	6.678	5.911	-0.767
36	6.468	5.780	-0.688
37	5.131	5.587	0.456
38	5.585	5.971	0.386
39	4.730	5.734	1.004
40	5.585	5.702	0.117
41	5.886	5.474	-0.412
42	5.284	5.274	-0.010
43	6.000	5.423	-0.578
44	5.602	5.781	0.179
45	4.876	5.348	0.472

46	6.319	5.873	-0.446
47	6.215	5.547	-0.668
48	6.051	5.364	-0.687
49	4.445	5.014	0.569
50	4.958	5.205	0.247
51	4.086	4.314	0.228
52	4.217	5.350	1.133
53	6.398	5.868	-0.530
54	5.569	6.338	0.769
55	4.663	4.356	-0.307
56	6.229	6.271	0.042
57	4.182	4.633	0.451
58	5.056	4.662	-0.394
59	5.009	5.048	0.039
60	5.149	5.202	0.053
61	5.886	5.417	-0.469
62	5.886	5.751	-0.135
63	4.801	5.212	0.411
64	5.201	6.210	1.009
65	6.959	6.406	-0.553
66	5.538	6.109	0.571
67	6.482	6.435	-0.047
68	5.921	5.929	0.008
69	6.769	6.326	-0.444
70	5.569	6.317	0.748
71	6.824	6.031	-0.793
72	6.051	6.131	0.080
73	7.222	6.133	-1.089
74	6.000	5.666	-0.335
75	6.620	6.532	-0.088
76	6.638	6.272	-0.366
77	5.495	5.964	0.469
78	5.959	6.508	0.549
79	6.181	6.435	0.254
80	5.187	5.216	0.029
81	7.301	6.943	-0.358
82	6.921	6.926	0.005
83	6.201	6.571	0.370

ARTIGO 2**Construction of 4D-QSAR models for use in the design of novel N-miristoiltransferase inhibitors of *Leishmania donovani*.**

Artigo redigido em concordância com as normas do periódico “*Journal of the Brazilian Chemical Society*”

(Versão preliminar)

Letícia Santos-Garcia^a, Daniela R. Silva^a, Letícia C. Assis^a, Tamiris M. de Assis^a, Giovanna C. Gajo^a, Teodorico C. Ramalho^a, Elaine F. F. da Cunha^{1*}

^aDepartment of Chemistry, Federal University of Lavras, P.O. Box 3037, 37200-000, Lavras, MG, Brazil

* elaine_cunha@dqi.ufla.br

Abstract

Protein N-myristoylation is catalysed by N-myristoyltransferase (NMT), an essential target in *Leishmania donovani*, the causative agent of kala-azar. Four-dimensional quantitative structure–activity relationship (4D-QSAR) analysis was applied to a series of 77 *Leishmania donovani* NMT inhibitors. Then, three new compounds were proposed using QSAR models. In addition, molecular docking was performed to investigate the binding affinities and interaction modes among the proposed compounds and the NMT active site. An ADME evaluation was performed and potential inhibitors demonstrated satisfactory pharmacokinetic properties.

Keywords: leishmaniasis; inhibitors; 4D-QSAR; docking; ADME

Introduction

Visceral leishmaniasis (VL), also known as kala-azar, is a parasitic infection caused by *Leishmania donovani* and if untreated, it is fatal in more than 95% of cases within 2 years after disease onset. In 2016, the World Health Organization (WHO) and Gilead Sciences reported that there are around 200,000 to 400,000 new cases worldwide, and the disease is endemic in more than 80 countries. VL is characterized by irregular bouts of fever, weight loss, enlargement of the spleen and liver, and anaemia¹.

Chemotherapy is the main way to deal with this disease. The first-line drug, pentavalent antimony, has several adverse effects, such as vomiting, nausea, anorexia, myalgia, abdominal pain, headache, arthralgia, and lethargy and can rarely cause the severe reaction of fatal cardiac arrhythmia². Other drugs, such as pentamidine, amphotericin B, paromomycin and miltefosine have several disadvantages, such as toxicity, lengthy treatment, need for hospitalization and drug resistance. Only one of the commonly used drugs, miltefosine, can be administered orally, however, it has a high teratogenic potential³. Furthermore, the high treatment cost is also a barrier to effective disease control in poor countries.

N-myristoyltransferase (NMT) catalyzes the covalent co-translational attachment of the fatty acid, myristate (C14:0) to the amino-terminal glycine residue after the removal of the initiating methionine⁴. N-Myristoylation plays an important role in protein-protein interactions of membrane proteins which, in turn, facilitate a variety of signal transduction pathways⁴. NMT has been found to be essential in *Leishmania donovani* life cycle⁵, becoming highly promising as a drug target for the treatment of VL.

In the present paper, the quantitative structure activity relationship in four dimensions (4D-QSAR) of 77 NMT inhibitors of *Leishmania donovani* developed by Leatherbarrow⁶ and coworkers to propose structural changes in these compounds in order to make them more

potent. Subsequently, a docking study was carried out to understand the interaction mode of the proposed compounds in the NMT active site. In addition, an ADME (absorption, distribution, metabolism and excretion) evaluation was performed to provide a notion of the toxicity of the proposed inhibitors.

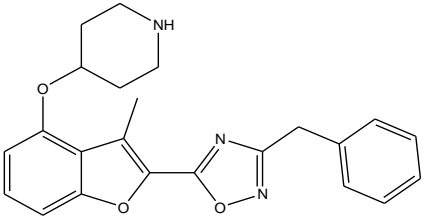
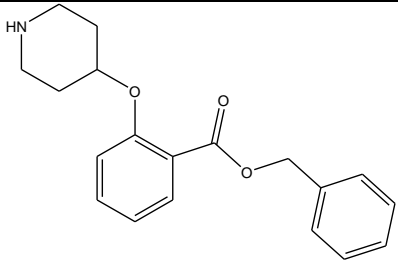
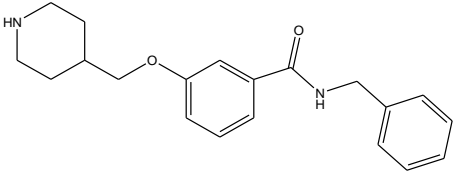
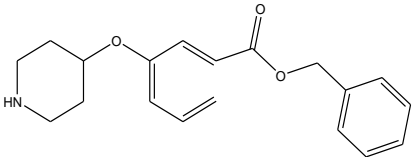
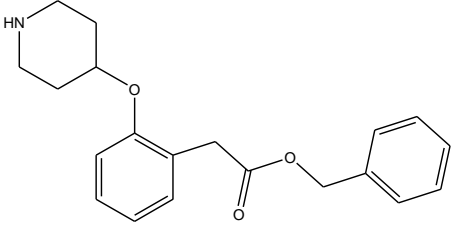
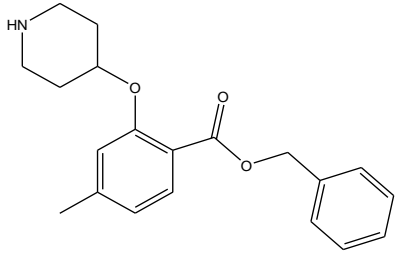
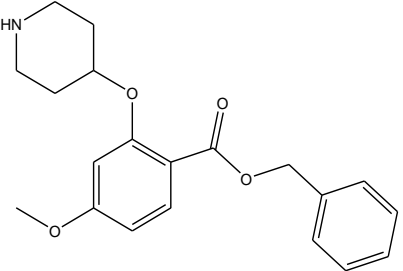
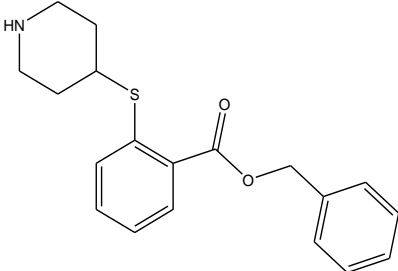
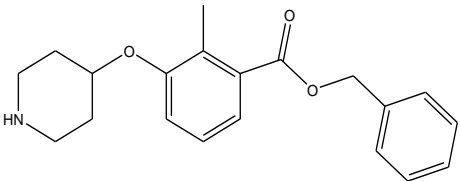
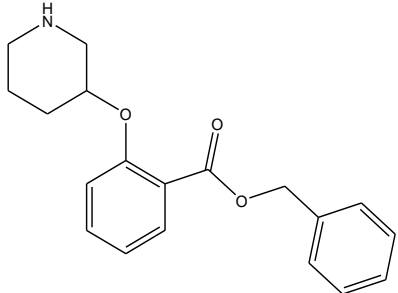
Experimental

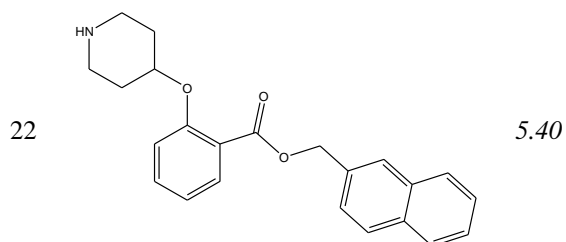
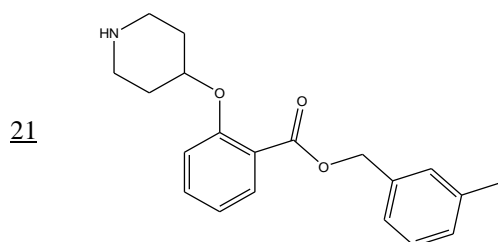
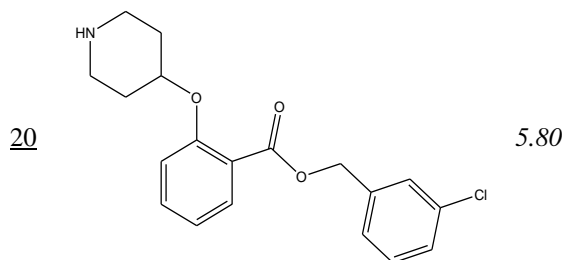
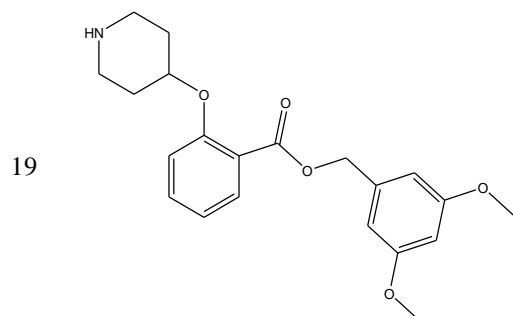
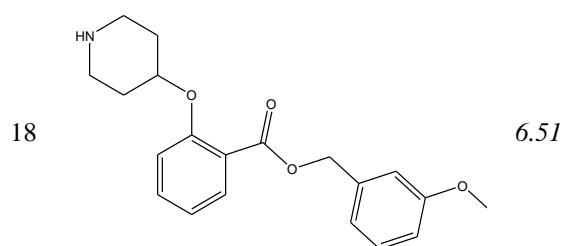
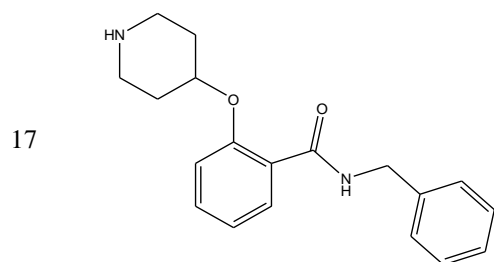
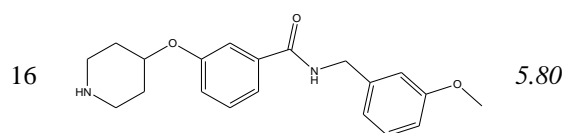
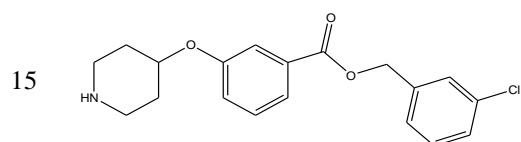
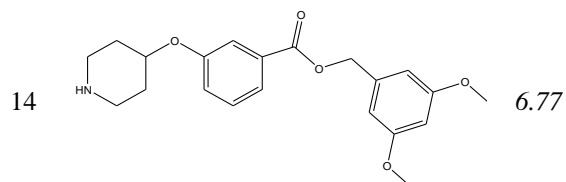
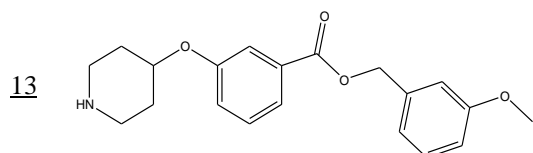
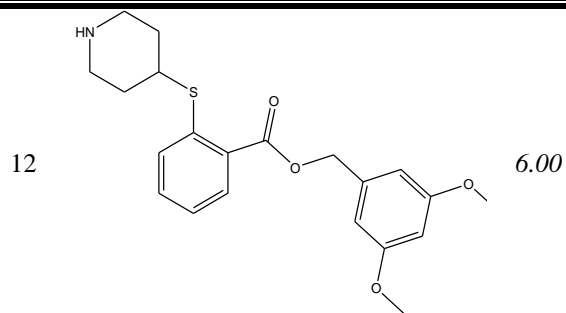
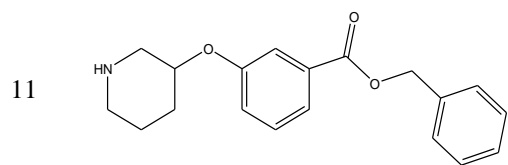
Biological Data

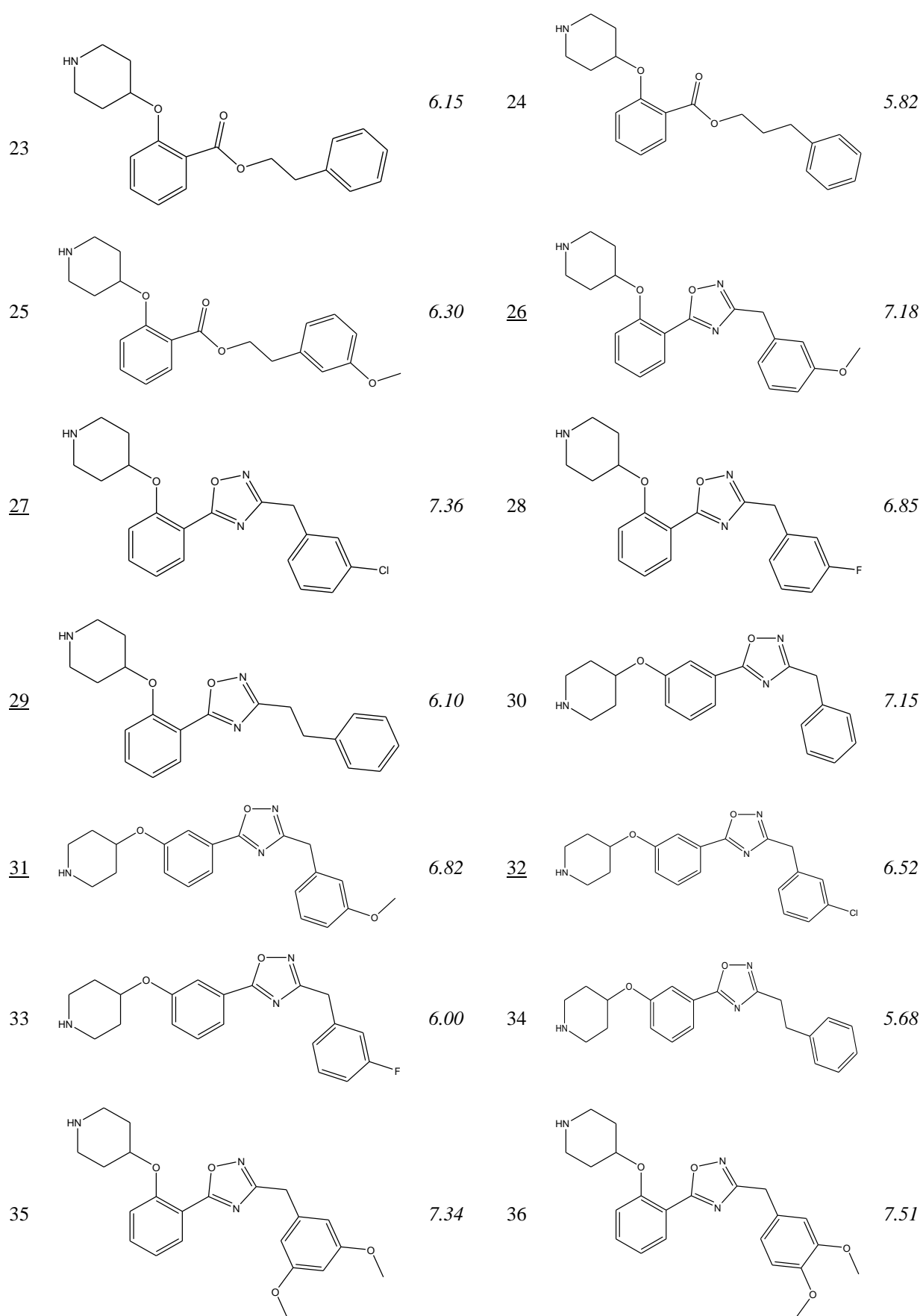
The 77 *Leishmania donovani* inhibitors studied were synthesized and pharmacologically tested by Leatherbarrow et al⁶. Among these compounds, 19 represent the test set (external validation). 3 test sets were formed to obtain greater reliability for the model. The first test group (Test set I) is formed by Molecules: **3; 6; 9; 13; 20; 21; 26; 27; 29; 31; 32; 41; 46; 55; 60; 63; 66; 71** and **72**, the second test group (Test set II) by Molecules: **7; 10; 16; 19; 24; 26; 33; 37; 40; 44; 50; 55; 56; 63; 64; 68; 71; 74** and **75**, and the third test group (Test set III) by Molecules: **11; 18; 19; 23; 24; 26; 28; 29; 30; 35; 37; 39; 40; 41; 46; 47; 61; 63** and **75**. The remaining compounds represent the Training set and are used for the construction of QSAR-4D models.

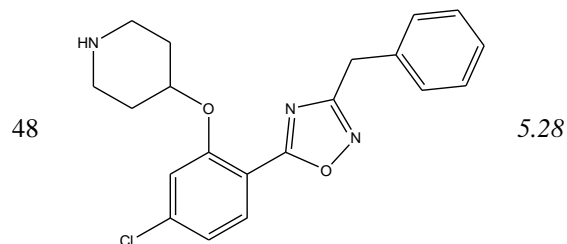
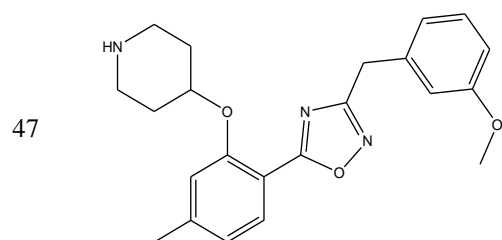
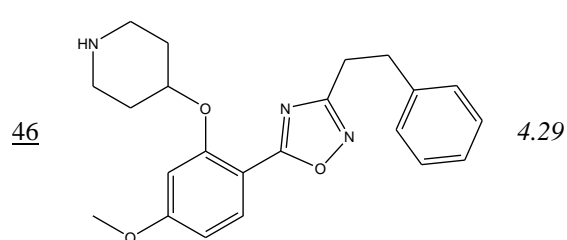
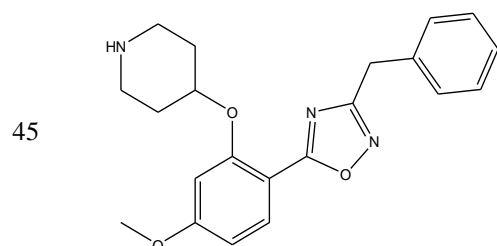
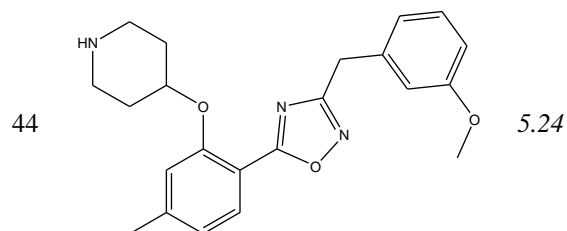
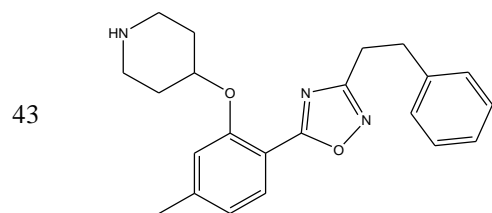
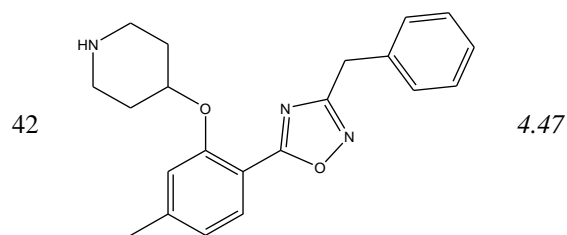
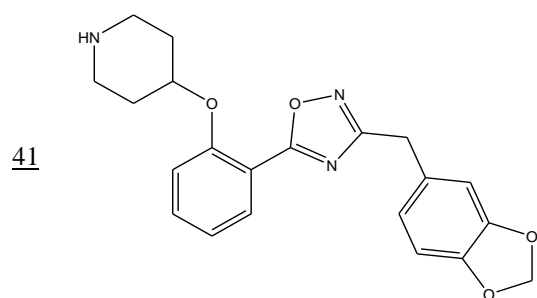
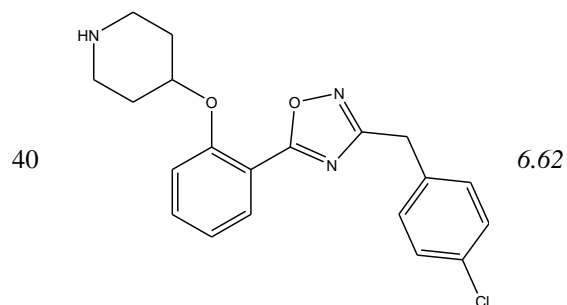
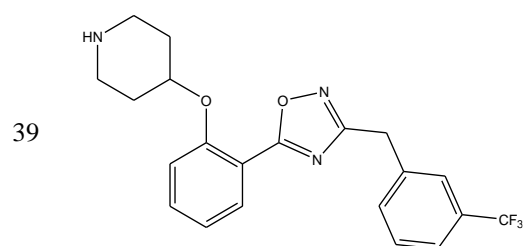
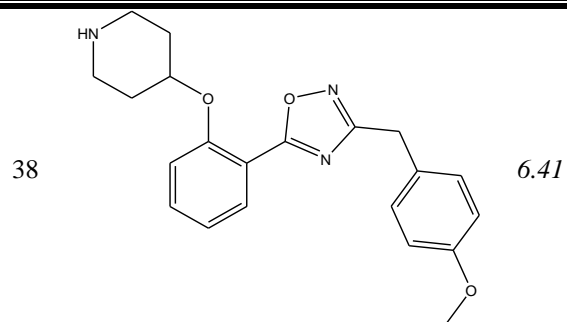
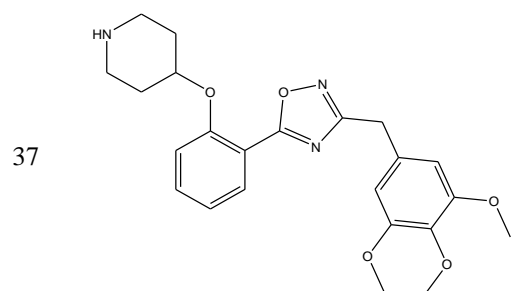
The biological activities of these compounds were reported as the concentration capable of inhibiting 50% of the enzyme activity (IC_{50}), measured using an adapted version of the sensitive fluorescence-based assay based on CoA detection by 7-diethylamino-3-(4-maleimido-phenyl)-4-methylcoumarin⁶. In addition, all pharmacological data were obtained from the same laboratory, eliminating the potential noise that might have been introduced by the pooling of data sets from different sources. The IC_{50} (μM) values were converted into molar units and then expressed in negative logarithmic units (pIC_{50}), and are represented in Fig.1. The range of pIC_{50} values for the training and test set spans at least four orders of magnitude (3.87 to 8.00), and the biological activity values show a regular distribution over the whole range.

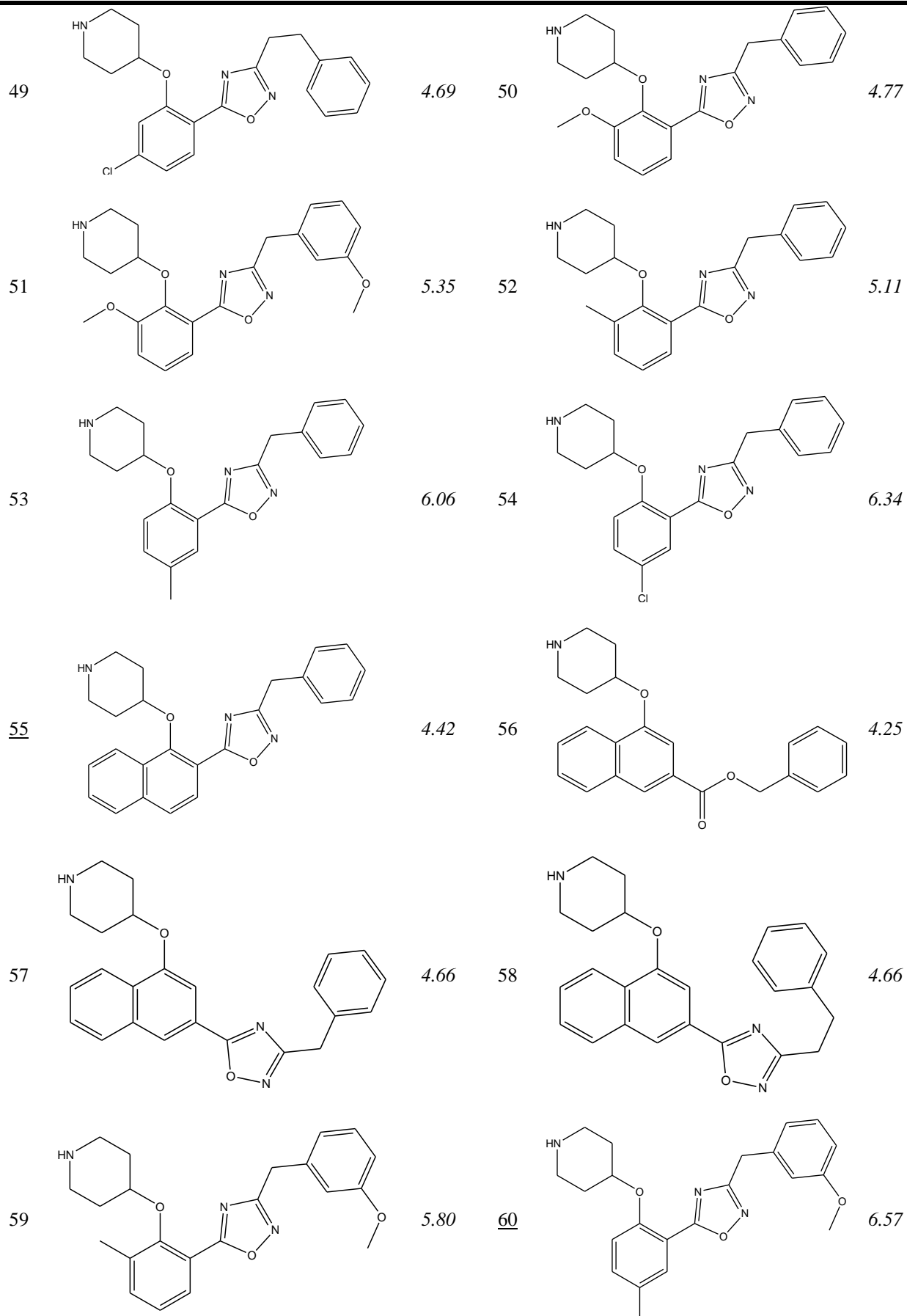
The three-dimensional (3D) models of the 77 compounds, Fig. 1, were constructed using the *HyperChem 7.0* software.⁷ The structures were geometry-optimized in vacuum, without any restriction, using the MM+ molecular mechanics force field (HyperChem), and latterly applying the semi-empirical RM1 Hamiltonian⁸, in order to assign the partial atomic charges.

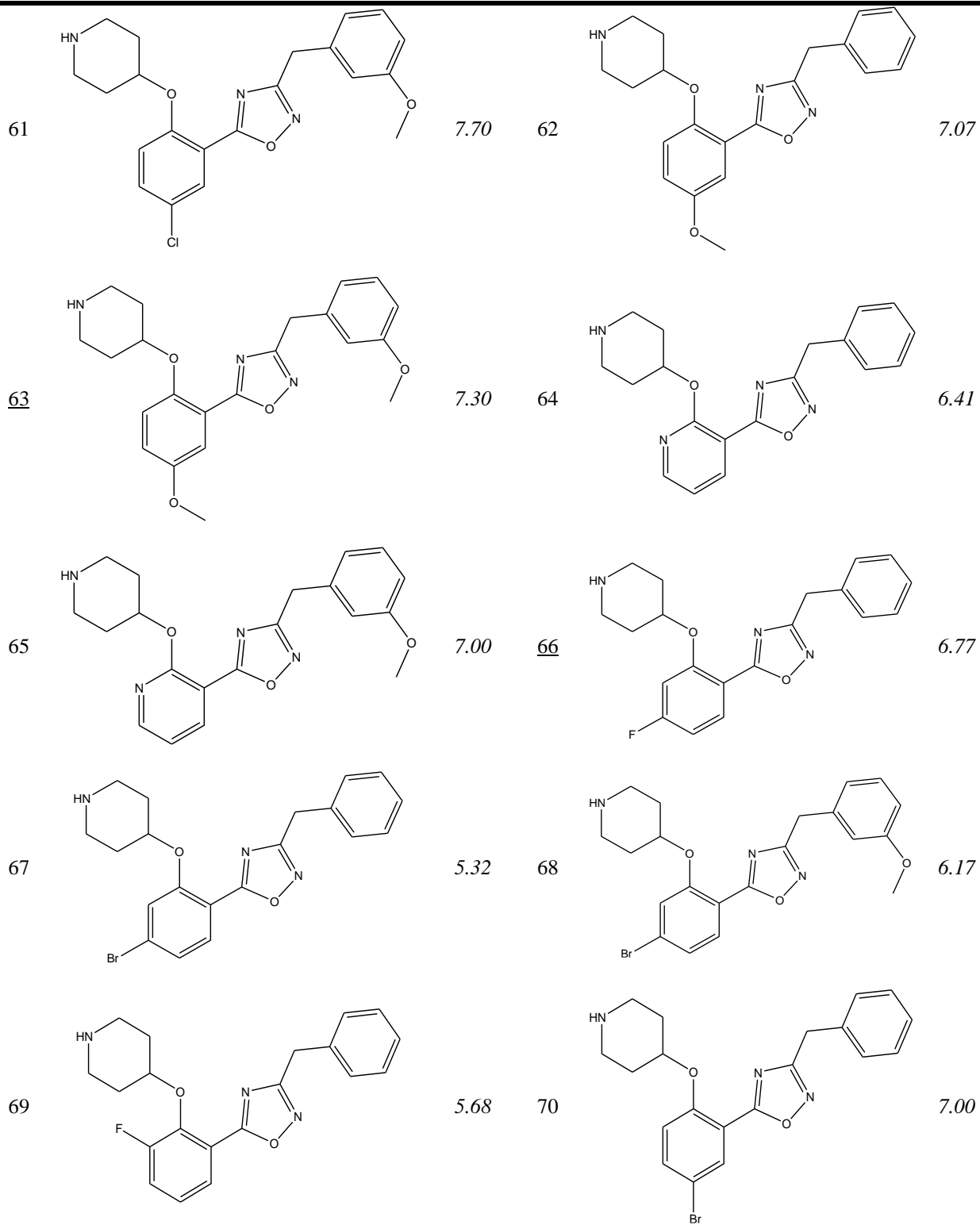
N°	Structure	pIC ₅₀	N°	Structure	pIC ₅₀
1		6.30	2		6.22
3		5.46	4		6.10
5		5.38	6		5.03
7		4.63	8		5.37
9		5.80	10		5.15











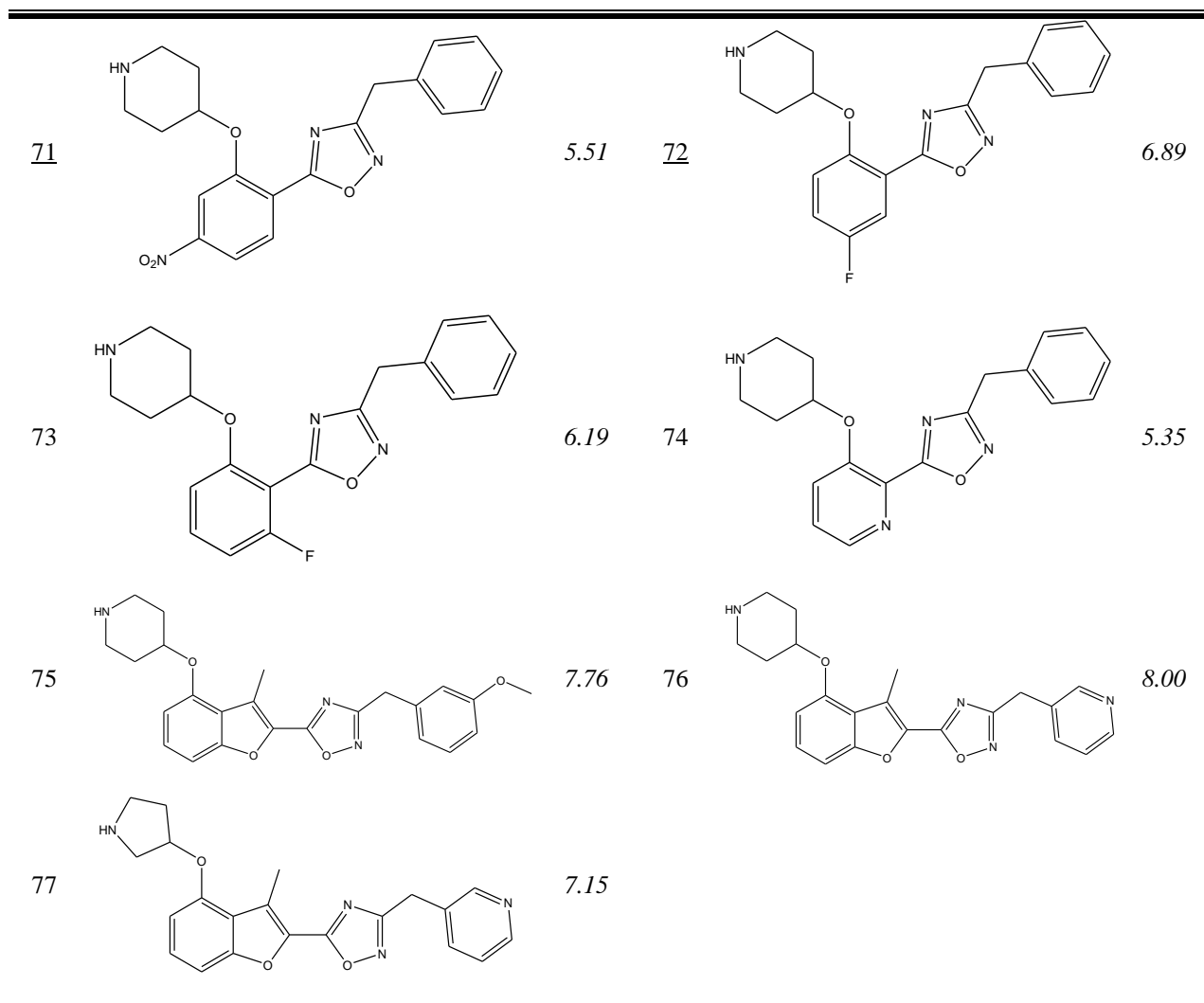


Figure 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC₅₀ values⁶. Compound numbers of Test set I are underlined.

Molecular dynamic simulation (MDS)

Molecular dynamics simulation (MDS) was carried out using the MOLSIM package in the 4D-QSAR program⁹, starting from the RM1 structures, in order to build the conformational ensemble profile (CEP) of each molecule. The temperature for MDS was set at 300 K, close to the temperature assays, with a simulation sampling time of 100 ps, and intervals of 0.001 ps. Thus, a total sample of 100,000 conformations of each ligand was produced. The MDS calculations were carried out employing a distance-dependent dielectric function, $\epsilon_r = D^*r_{ij}$,

which was set to $3*r_{ij}$ in order to try to model the solvent effect.

Alignment definition

In this study, we will assume that all molecules bind to the enzyme in a similar way, since the compounds are structural analogues. In general, the alignments are chosen to span the common framework of the molecules in the training and test sets. Seven alignments were performed using atoms of the benzene ring. Three-ordered atom trial alignments were selected: (1) A-B-C, (2) A-B-D (3) C-D-E, (4) C-D-F, (5) B-A-C, (6) C-B-A and (7) D-A-F, using Compound **76** as a reference (Fig. 2). The conformational ensemble profile (CEP) for each compound was obtained after the MDS step was overlaid onto a cubic lattice with grid cell size of 1Å.

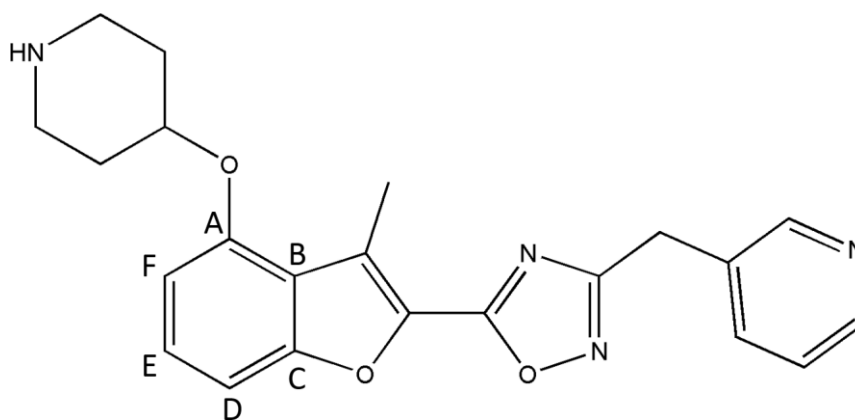


Figure 2. Ordered atom letter codes (A, B, C, D, E and F) used in the 4D-QSAR analysis defines the three trial alignments. Compound **76** ($pIC_{50} = 8.00$) is used to define the atom letter code.

Interaction pharmacophore elements

Each atom was classified into seven types of interaction pharmacophore elements (IPE):

i) any type (any); ii) nonpolar (np); iii) polar-positive charge density (p+); iv) polar-negative

charge density (p-); v) hydrogen bond acceptor (hba); vi) hydrogen bond donor (hbd); and vii) aromatic systems (ar).¹⁰ The use of IPEs allows each of the compounds in a training set to be partitioned into sets of structure types and/or classes with respect to possible interactions with a common receptor. The occupancy of the grid cells by each IPE type are recorded over the conformational assembly profile, and forms the set of grid cell occupancy descriptors (GCOD) to be utilized as the pool of trial descriptors in the model building and optimization process.¹⁰

4D-QSAR model calculation

4D-QSAR analysis generates an enormous number of QSAR descriptors because of the large number of grid cells and the number of IPEs.¹⁰⁻¹¹ In order to exclude data noise from databases generated by the alignments, partial least-squares (PLS) regression analysis was performed as a data reduction fit between the observed dependent variable measures and the corresponding set of GCOD values. Furthermore, PLS identifies the most highly weighted GCODs from data set of local grid cells.¹⁰

The two hundred GCODs with the highest weight from the data reduction were chosen optimized using a combined genetic function approximation (GFA) approach,¹² implemented in the 4D-QSAR program.¹³ Optimizations were initiated using 100 randomly generated models and 10,000-100,000 crossover operations. Mutation probability over the crossover optimization cycle was set at 10-30%. The smoothing factor, the variable that specifies the number of descriptors in the QSAR models, was varied between 1.0 and 3.0, in order to obtain equations with no more than eleven terms.

Model validation

Validation is an important factor in QSAR modeling. The best models, resulting from the 4D-QSAR study were based on different criteria.¹⁴⁻¹⁸

- i) Coefficient of determination (r^2): is a measure of how well the regression line represents the data.
- ii) Leave-one-out cross-validation (LOOCv) correlation coefficient (q^2): estimating the performance of a predictive model.
- iii) Adjusted cross-validated squared correlation coefficient (q^2_{adj}): allows the comparison between models with different number of variables.
- iv) Least-square error of fit (LSE);
- v) Friedman's lack-of-fit (LOF): estimates the quality of a model. This measure penalizes appropriately for the addition of terms to the equation (and consequent loss of degrees of freedom) in such a way to resist over-fitting.
- vi) Correlation coefficient of external validation set (R^2_{pred}): reflects the degree of correlation between the observed ($Y_{Exp(test)}$) and predicted ($Y_{Pred(test)}$) activity data of the test set:

$$R^2_{Pred} = 1 - \frac{\sum_1^n (Y_{Exp(test)} - Y_{Pred(test)})^2}{\sum_1^n (Y_{Exp(test)} - \bar{Y}_{training})^2} \quad \text{Eq.1}$$

Where $\bar{Y}_{training}$ is average value for the dependent variable for the training set.

- vii) Modified r^2 ($r^2_{m(test)}$) equation determining the proximity between the observed and predicted values with the zero axis intersection:

$$r^2_m = r^2 \left(1 - \left| \sqrt{r^2 - r_0^2} \right| \right) \quad \text{Eq.2}$$

- viii) Y-Randomization (R^2_r): consists of the random exchange of the independent variable values. Thus, the R^2_r value must be less than the correlation coefficient of the non-randomized models.

- ix) R^2_p : penalizes the model R^2 for the difference between the squared mean correlation coefficient (R^2_r) of randomized models and the square correlation coefficient (R^2) of the non-randomized model:

$$r_p^2 = r^2 * \sqrt{r^2 - r_r^2} \quad \text{Eq.3}$$

Docking studies

The crystal structural coordinates of NMT at 1.42Å resolution was obtained from the Protein Data Bank (PDB code: 2WUU),⁵ and the protein has been crystallized in complex with the non-hydrolysable substrate analogue S-(2-oxo)pentadecyl-CoA. For the docking procedure the Molegro Virtual Docker (MVD) program was used.¹⁹ Molecular docking uses the receptor structure as a template for the development of new ligands, estimating binding affinity between linker and receptor. The MolDock scoring function (E_{Score}) used in MVD, is defined by:

$$E_{\text{Score}} = E_{\text{Inter}} + E_{\text{Intra}} \quad \text{Eq.4}$$

wherein E_{Inter} is the ligand–protein interaction energy and E_{Intra} is the internal energy of the ligand.

Conformational selection.

In the 4D-QSAR method, the conformation of each compound can be postulated as the lowest-energy conformer state from the set sampled for each compound, which predicted the maximum activity using the optimum 4D-QSAR model.¹⁰

Results and Discussion

Alignments evaluation

In this study seven alignments were evaluated for each of the three test groups. The GA-PLS analysis using grid cells of 1.0 Å generated several models or equations. The statistical parameters of seven alignments studied for Test set I (Compounds **3**; **6**; **9**; **13**; **20**; **21**; **26**; **27**; **29**; **31**; **32**; **41**; **46**; **55**; **60**; **63**; **66**; **71** and **72**), Test set II (Compounds **7**; **10**; **16**; **19**; **24**; **26**; **33**; **37**; **40**; **44**; **50**; **55**; **56**; **63**; **64**; **68**; **71**; **74** and **75**) and Test set III (Compounds **11**; **18**; **19**; **23**; **24**; **26**; **28**; **29**; **30**; **35**; **37**; **39**; **40**; **41**; **46**; **47**; **61**; **63** and **75**) are shown in Table 1, Table 2 and Table 3, respectively.

Table 1. Statistical parameters evaluated in the 4D-QSAR analysis for the seven performed alignments of the Test set I.

Alignment	r^2	q^2	q^2_{adj}	LSE	LOF	R^2_{Pred}	R^2_{m} (test)	R^2_{r}	R^2_{p}	GCODs
A1	0.76	0.70	0.65	0.21	0.39	0.74	0.54	0.11	0.61	5
A2	0.73	0.66	0.61	0.22	0.49	0.82	0.48	0.09	0.68	5
A3	0.74	0.69	0.64	0.24	0.52	0.67	0.45	0.08	0.70	5
A4	0.69	0.69	0.64	0.24	0.52	0.67	0.45	0.09	0.70	5
A5	0.72	0.66	0.61	0.24	0.72	0.73	0.51	0.08	0.68	5
A6	0.74	0.68	0.63	0.23	0.41	0.75	0.45	0.07	0.71	5
A7	0.71	0.64	0.59	0.24	0.53	0.77	0.58	0.10	0.66	5

Table 2. Statistical parameters evaluated in the 4D-QSAR analysis for the seven performed alignments of the Test set II.

Alignment	r^2	q^2	q^2_{adj}	LSE	LOF	R^2_{Pred}	R^2_m (test)	R^2_r	R^2_p	GCODs
<i>B1</i>	0.74	0.69	0.64	0.20	0.37	0.69	0.52	0.08	0.71	5
<i>B2</i>	0.72	0.66	0.61	0.23	0.50	0.78	0.55	0.08	0.68	5
<i>B3</i>	0.70	0.63	0.59	0.24	0.53	0.77	0.55	0.08	0.65	5
<i>B4</i>	0.64	0.57	0.54	0.28	0.59	0.76	0.44	0.08	0.60	4
<i>B5</i>	0.72	0.66	0.61	0.22	0.49	0.78	0.58	0.08	0.68	5
<i>B6</i>	0.70	0.64	0.59	0.22	0.41	0.83	0.63	0.08	0.66	5
<i>B7</i>	0.73	0.67	0.62	0.24	0.52	0.69	0.49	0.08	0.69	5

Table 3. Statistical parameters evaluated in the 4D-QSAR analysis for the seven performed alignments of the Test set III.

Alignment	r^2	q^2	q^2_{adj}	LSE	LOF	R^2_{Pred}	R^2_m (test)	R^2_r	R^2_p	GCODs
<i>C1</i>	0.74	0.68	0.63	0.21	0.38	0.74	0.57	0.07	0.70	5
<i>C2</i>	0.72	0.65	0.60	0.23	0.50	0.75	0.52	0.08	0.67	5
<i>C3</i>	0.71	0.64	0.59	0.25	0.60	0.65	0.33	0.11	0.65	5
<i>C4</i>	0.71	0.65	0.60	0.23	0.51	0.66	0.37	0.12	0.65	5
<i>C5</i>	0.73	0.66	0.61	0.22	0.49	0.72	0.46	0.09	0.68	5
<i>C6</i>	0.75	0.69	0.64	0.22	0.49	0.64	0.51	0.08	0.70	5
<i>C7</i>	0.72	0.65	0.60	0.24	0.52	0.73	0.56	0.08	0.67	5

All tested alignments showed q^2 and q^2_{adj} values higher than 0.5. This reveals that the model can be a useful tool for predicting affinities of new compounds based on these structures; r^2 greater than 0.7 indicates that the model is correlated and may be considered to represent the training set in the same manner.¹⁰ Thus, the alignments A4 and B4 were eliminated from the analysis because they had r^2 lower than 0.7.

Regarding external validation, R^2_{pred} values for all alignments were greater than 0.5. Analyzing $R^2_{\text{m(test)}}$ values, alignments A2, A3, A4, A6, B4, B7, C3, C4 and C5 were excluded, since they should be higher than 0.5.

Another parameter used in the validation of the models was R^2_{p} and all values are greater than 0.5, which means that R^2_{p} values are acceptable for a good QSAR model. In addition, all R^2_{r} values are well below those of r^2 . As alignment A1 from Test set I showed the highest r^2 , q^2 and q^2_{adj} values, it was chosen as the best alignment in this 4D-QSAR study. In addition, Alignment A1 presents good external validation values. We will only present the analysis of the best model derived from A1.

Equation 5 presents Model A1. The statistical measures, including the values of r^2 , q^2 , q^2_{adj} , LSE, LOF, R^2_{pred} , $R^2_{\text{m(test)}}$, R^2_{p} and R^2_{r} , are presented below. Each GCOD is labeled as “x, y, z, IPE” which represent the cartesian coordinate positions of the selected grid cell (x, y, z) and the respective atom type (IPE).

Model A1 – Eq.5:

$$\begin{aligned} \text{pIC}_{50} = & 5.76 + 1.19(0, -2, 4, \text{np}) - 0.56(0, 3, 2, \text{any}) + 1.46(3, -2, -1, \text{np}) \\ & + 2.03(0, -3, 2, \text{any}) - 0.95(0, 4, -1, \text{any}) \end{aligned}$$

$n = 58$, GCODs = 5, $r^2 = 0.76$, $q^2 = 0.70$, $q^2_{\text{adj}} = 0.65$, LSE = 0.21, LOF = 0.39, $R^2_{\text{pred}} = 0.74$, $R^2_{\text{m(test)}} = 0.54$, $R^2_{\text{r}} = 0.11$, $R^2_{\text{p}} = 0.61$.

Model A1 generated five descriptors, where GCODs (0,-2,4,**np**), (3,-2,-1,**np**), (0,-3,2,**any**), present positive coefficients (Eq. 5) and correspond to favorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Thus, substituents in these positions increase the effectiveness of the compounds. GCODs (0,3,2,**any**) and (0,4,-1,**any**) have negative coefficients and correspond to unfavorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Thus, the occupation of GCODs (0,3,2,**any**) and (0,4,-1,**any**) decreases the compound potency.

The cross-correlation matrix of the GCODs from Model A1 (Eq. 5, Table 4) was computed in order to determine if two or more highly correlated GCODs appear in the same 4D-QSAR model. We can observe that there is no correlation ($r > 0.7$) between the GCODs. Therefore, each descriptor contributes in different ways to the 4D-QSAR models.²⁰

Table 4. Cross-correlation matrix of the GCODs of Model A1 obtained in the 4D-QSAR analysis.

	(0,-2,4,1)	(0,3,2,0)	(3,-2,-1,1)	(0,-3,2,0)	(0,4,-1,0)
(0,-2,4,1)	1				
(0,3,2,0)	0.245	1			
(3,-2,-1,1)	-0.403	-0.416	1		
(0,-3,2,0)	-0.324	0.069	0.098	1	
(0,4,-1,0)	-0.27	0.222	-0.037	-0.216	1

Figure 3 shows William's plot for the Model A1. We can observe that the compounds belong to the same chemical domain, thus validating the model.

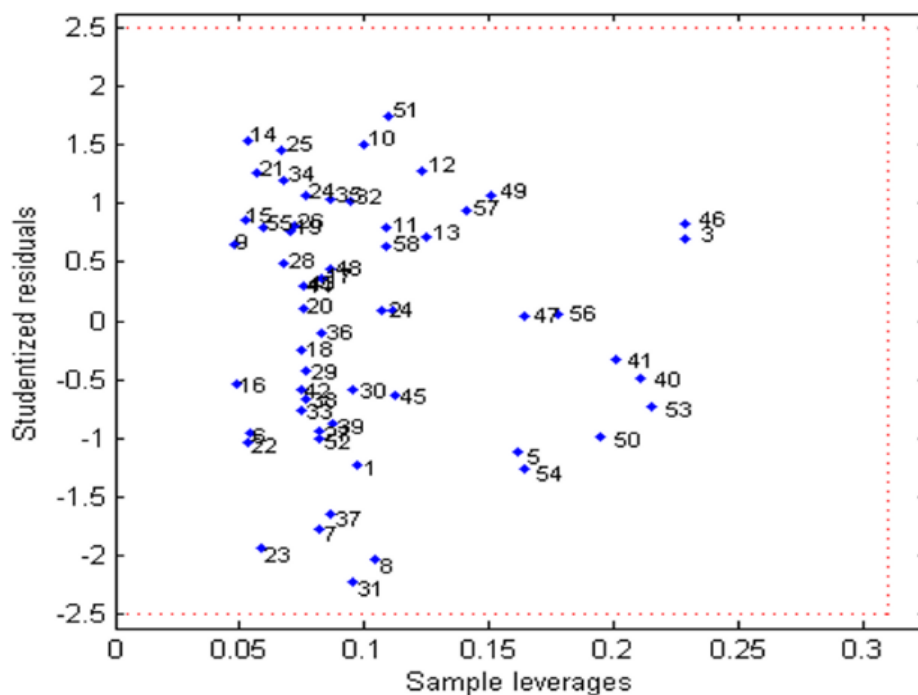


Figure 3. Plot of sample leverage versus Student residuals for the Model A1. A graphic representation of the descriptors of Model A1 is shown in Fig. 4 using Compound **76** as a reference. Light and dark spheres represent GCODs with positive and negative coefficients, respectively, in accordance with Eq.5 (Model A1).

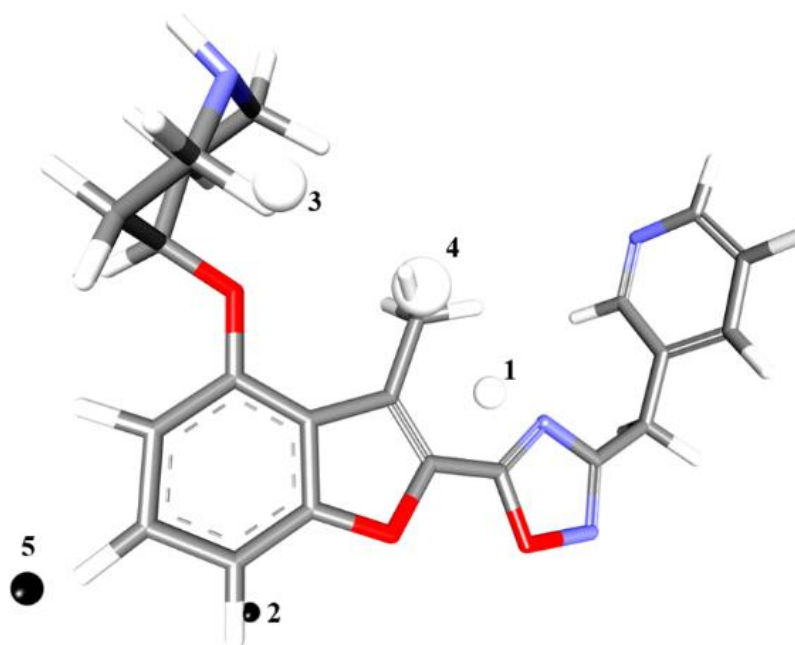


Figure 4. Graphic representation of Compound **76** according to the 4D-QSAR Model A1. GCODs occupancy represented by white spheres contribute to increasing the potency of compounds, and black spheres to decrease the potency of the compounds. The GCODs described are: (1) (0,-2,4,**np**), (2) (0,3,2,**any**), (3) (3,-2,-1,**np**), (4) (0,-3,2,**any**), (5) (0,4,-1,**any**).

GCOD-1 (0,-2,4,**np**), Fig. 5, has a positive coefficient of 1.19 and has a nonpolar IPE. This grid cell shows the highest frequency of occupation for compounds having a benzyl group at this site of the molecule, such as Molecules **2**, **11** and **15**. In contrast, compounds with only one phenyl group have no occupancy in this grid cell. Furthermore, the Molecules **61**, **75** and **76**, which are the most active in this series have no occupancy for this descriptor, which is favorable for the potency of the compounds.

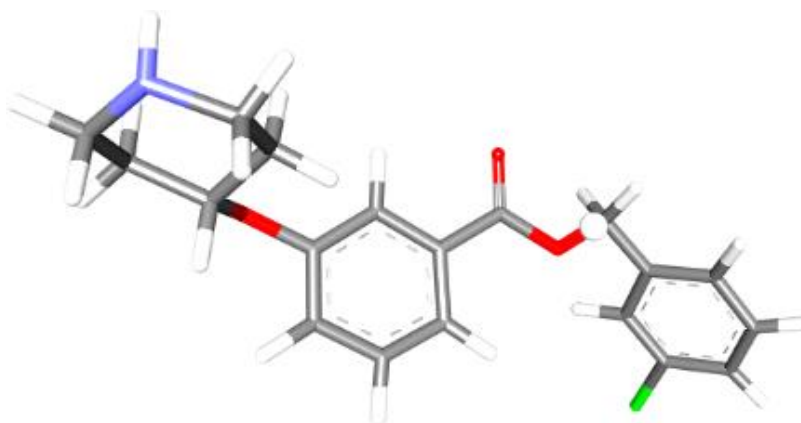


Figure 5. Representation of compound **15** and GCOD-1 (0,-2,4,**np**) (white sphere) obtained from Model A1.

GCOD-2 (0,3,2,**any**) (Fig.6) contributes to decrease compound potency and presents a coefficient of -0.56. This grid cell represents a non-specific IPE. The great majority of the compounds present a high GCOD-2 occupation frequency and have a hydrogen atom in this grid cell. However, Compounds **53**, **54**, **62** and **70** present low occupancy values for this negative GCOD and in this position have the following substitutions: CH₃, Cl, O-CH₃ and Br. Thus, the presence of a hydrogen atom in this grid cell results in a decrease in the potency of the compounds. Among the most active compounds, Molecules **75** and **76** have the hydrogen atom in this position, which is detrimental to their activity.

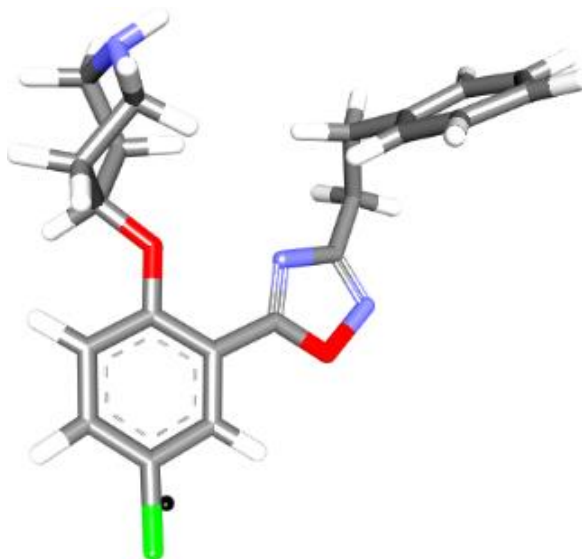


Figure 6. Representation of compound **54** and GCOD-2 (0,3,2,**any**) (black sphere) obtained from Model A1.

GCOD-3 (3,-2,-1,**np**) shows a positive coefficient and corresponds to a nonpolar type (IPE) (Fig. 7). The GCOD (3,-2,-1,**np**) is located near the hydrogen atom in the piperidine ring and has the highest frequency of occupation for Compounds **61**, **70**, **73** and **75**. Although most of the compounds present the piperidine ring in the same position, these compounds did not show high occupancy for this descriptor, because during the molecular dynamics simulation, they assumed a different conformation, as exemplified by Compound **17**.

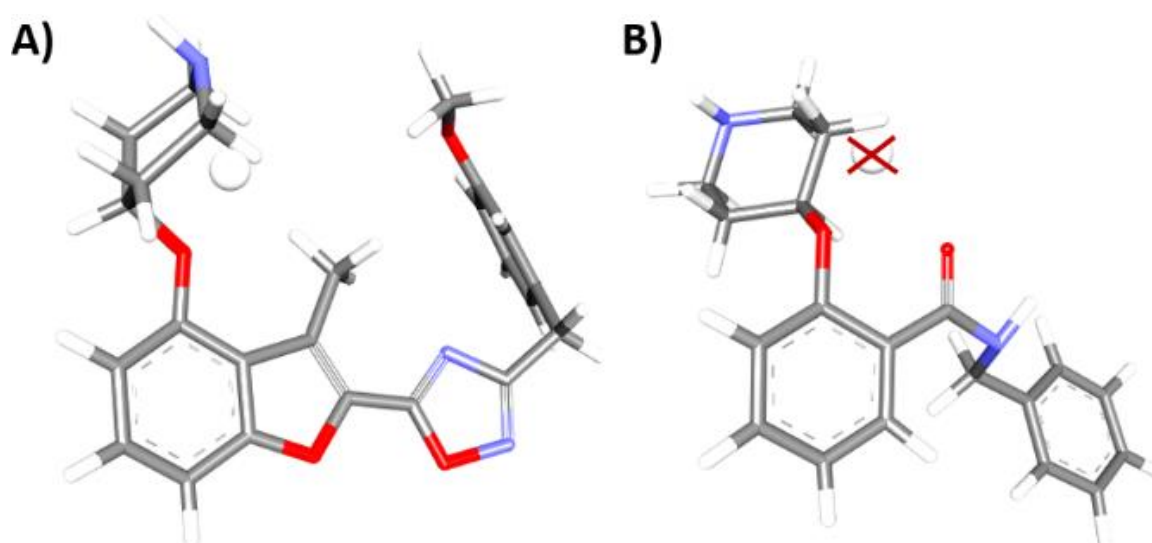


Figure 7. Representation of GCOD-3 (3,-2,-1,**np**) (white sphere) and **A)** compound **75**; **B)** Compound **17**.

The GCOD-4 (0,-3-,2,**any**) (Fig. 8), also has a positive coefficient of 2.03 and is the descriptor that contributes most to the increased effectiveness of the compounds. This grid cell represents a non-specific IPE and is located near the methyl group in benzofuran, 2,3-dihydro-3-methyl. This descriptor shows the highest occupation frequency for Compounds **1**, **75**, **76** and **77**, which have benzofuran, 2,3-dihydro-3-methyl in their structures. Thus, potential inhibitors would benefit from the exploitation of this region with methyl groups.

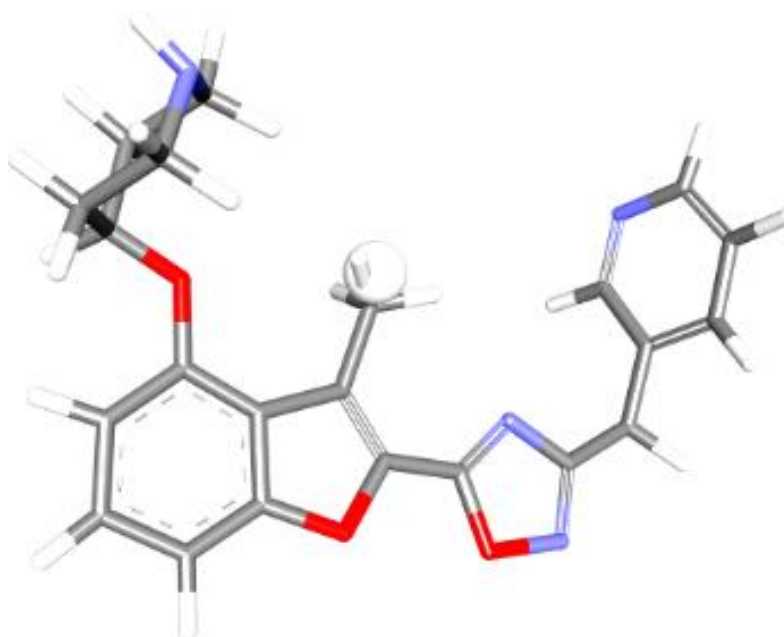


Figure 8. Representation of compound **76** and GCOD-4 (0,-3-,2,**any**) (white sphere) obtained from Model A1.

Finally, the GCOD-5 (0,4,-1,**any**) (Fig. 9) corresponds to a non-specific (IPE) and presents the most negative coefficient, impairing the activity of the compounds. Unlike GCOD-2, the substitution by CH₃, Cl, F, O-CH₃ and Br atoms or another benzene ring forming a naphthalene group in the GCOD-5, is detrimental to the potency of the compounds.

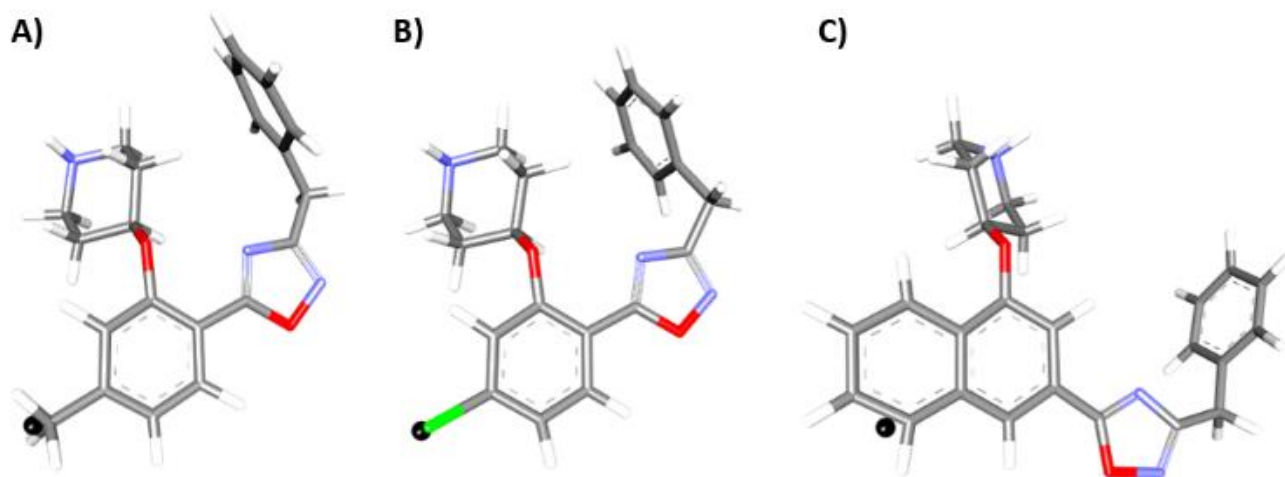


Figure 9. Representation of GCOD-5 (0,4,-1,**any**) (black sphere) and **A)** compound **42**; **B)** Compound **48**; **C)** Compound **57**.

Subsequently, Compounds **A**, **B** and **C** were proposed and their activities were predicted using Model A1. The predicted activity of Compound **76**, the most active in the series, was 7.57. The structure of the three compounds and their predicted pIC_{50} values are shown in Fig. 10 and all of them showed predicted activity values higher than Compound **76**.

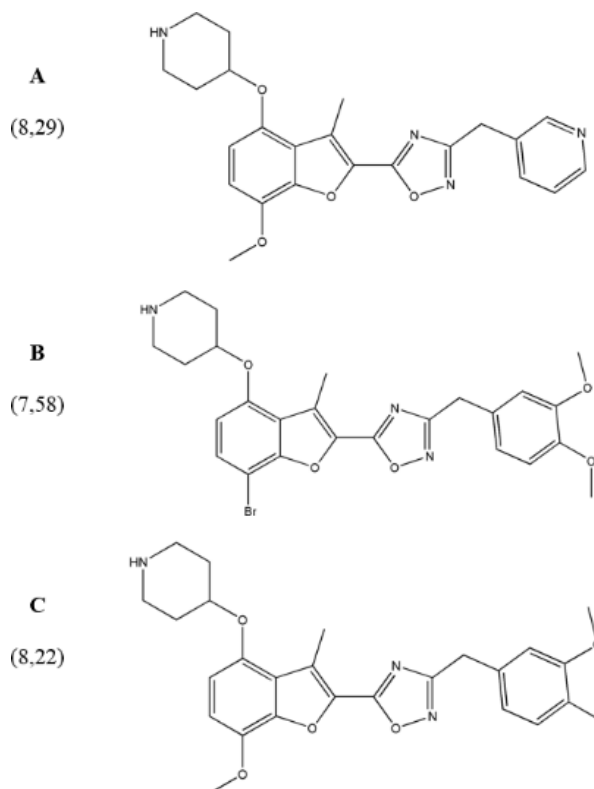


Figure 10. Proposed compounds using the Model A1 (predicted pIC_{50} values are shown in parenthesis).

To validate these predictions, docking studies were performed for proposed structures and also for Compound **76**. The potential binding sites of NMT were calculated and a cavity of 351.65 Å³ (surface = 1208.16Å²) was observed close to Tyr80, Val81, Glu82, Ser86, Met87, Phe88, Arg89, Phe90, Tyr92, Phe96, Asn167, Phe168, Thr203, Ala204, Gly205, Val206, Tyr217, Phe218, His219, Phe232, Tyr326, Ile328, Pro329, Ser330, Leu341, Ala343, Tyr345, Val346, Val374, Asn376, Met377, Val378, Ile380, Leu381, Asn383, Gly395, Asp396, Gly397, His398, Leu399, Tyr401, Val419, Met420, Leu421.

By analyzing the hydrogen bond formed between the compounds and the NMT active site, we observed: (i) the proposed Compounds **A** and **C** establish a hydrogen bond with: Tyr217, Tyr326 and Tyr345; however, Compounds **76** and **B** do not realize a hydrogen bond with the amino acid Tyr326. This portion of the molecule must be related to the GCOD-2, in which, the substitution of a hydrogen atom by different groups (CH₃, Br, Cl and O-CH₃) is favorable, as we have seen. The docking analysis also allows concluding that the presence of a group capable of hydrogen bonding with Tyr326 makes the compound more active. The structural difference between Compounds **A** and **76** is that Compound **A** has a group capable of hydrogen bonding with the amino acid Tyr326, while Compound **76** has a hydrogen atom, which is unfavorable. Thus, the activity value predicted by the Model A1 for Compound **A** is much higher than that of Compound **76**, as well as the interaction energy of the docking (Table 5). (ii) Compound **A** still makes a hydrogen bond with the amino acid Gly 397, near the pyridine. (iii) Compounds **B** and **C** interacted with Ser330 by a hydrogen bond. Fig. 11 shows the docking poses for the compounds analyzed inside the NMT active site and the interacting amino acid residues.

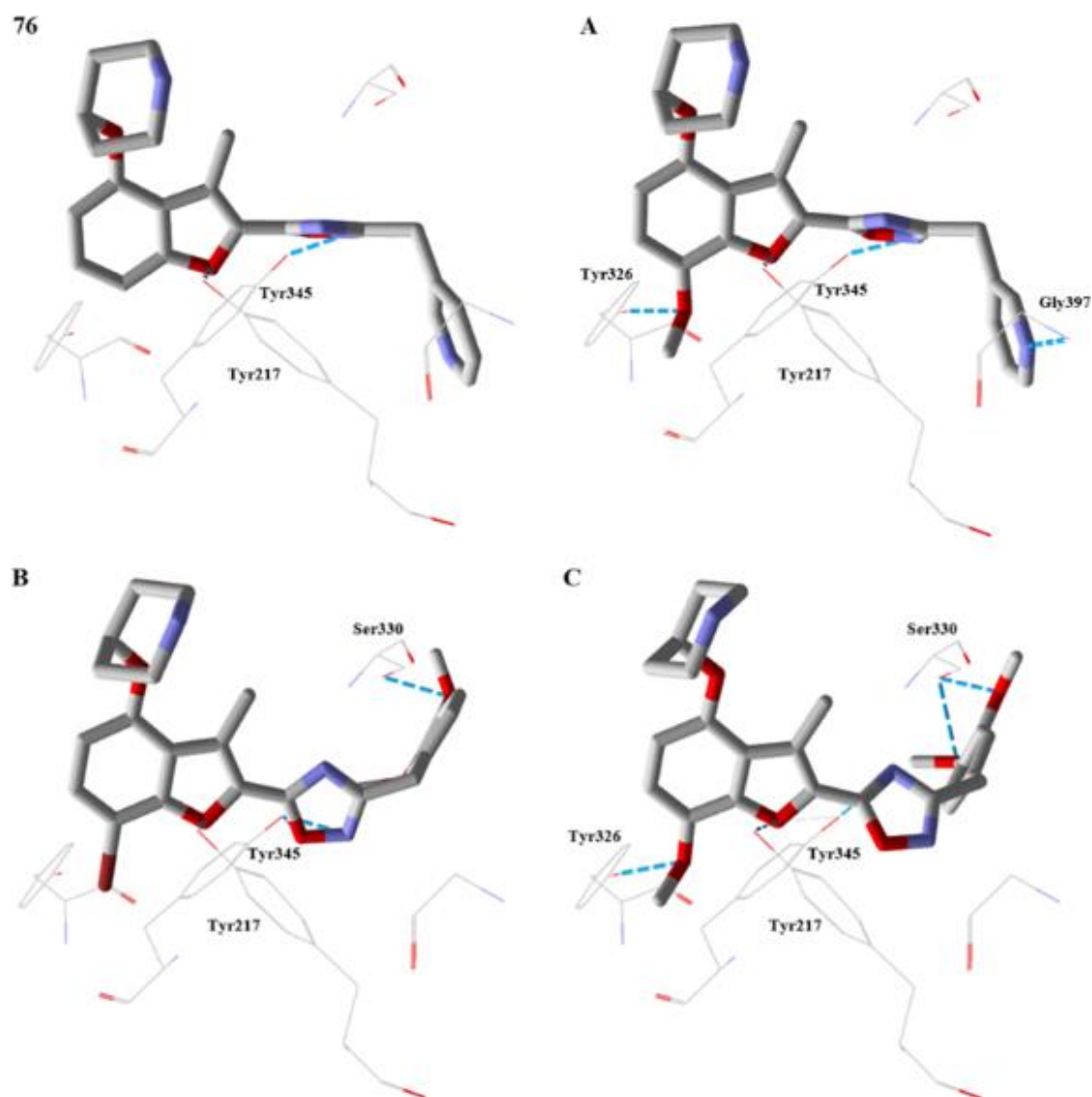


Figure 11. Docking poses for the compounds **76**, **A**, **B**, **C** and main interactions (hydrogen bonds).

Table 5. Predicted pIC_{50} , interaction energy (kcal mol^{-1}) values and binding hydrogen (kcal mol^{-1}) energy of compound/protein.

Compound	Predicted pIC_{50}	Interaction energy
76	7.57	-109.61
A	8.29	-126.19
B	7.58	-116.44
C	8.22	-127.32

In addition to the hydrogen bonding interactions, it was possible to note the presence of an apolar moiety which includes the amino acids Val81, Ile328 and Phe90 located near the methyl in benzofuran, 2,3-dihydro-3-methyl group (Fig. 12), as indicated by the GCOD-4. The presence of several aromatic amino acids in the active site, Fig. 13, as Tyr80, Phe88, Phe90, Tyr217, Phe232, Tyr326, Tyr345 and Tyr401 also contributes to the π - π staking interaction between these amino acids and aromatic moieties of inhibitors stabilizing them.

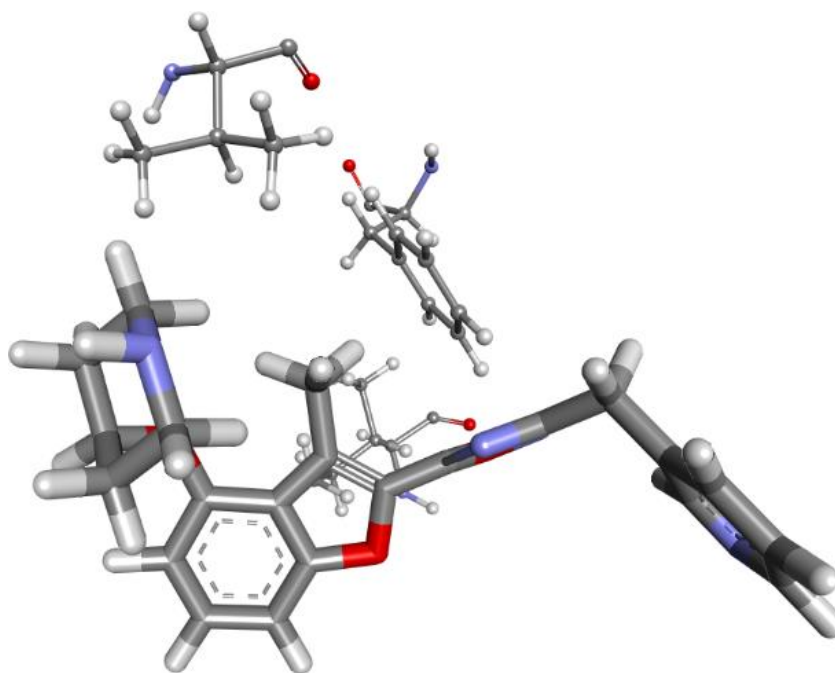


Figure 12. Docking pose for the compound **76** and apolar moiety next to the methyl group in the NMT active site.

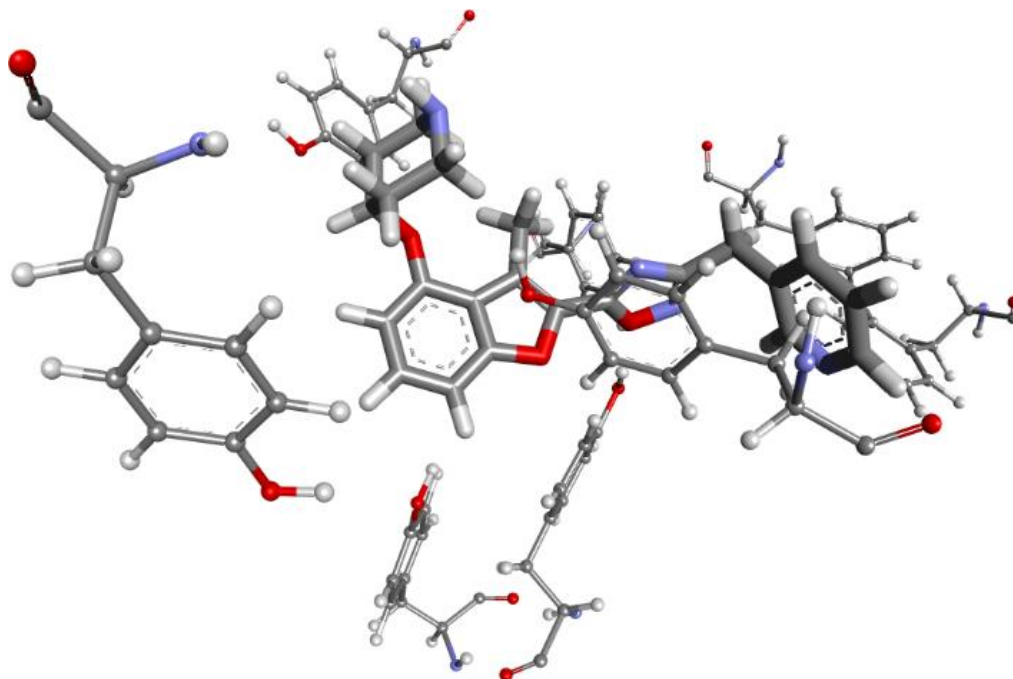


Figure 13. Docking pose for the compound **76** and aromatic amino acids in the NMT active site.

The absorption, distribution, metabolism and excretion (ADME) of a drug are important pharmacokinetic properties that must be met in the elaboration of a drug. One test for druglikeness is Lipinski's Rule of Five where molecular properties are closely related to the oral bioavailability of a drug,²¹ wherein compounds should not violate more than one rule. In this rule, the compounds should present $\log P \leq 5$, molecular weight ≤ 500 , number of hydrogen bond acceptors ($n_{\text{ON}} \leq 10$), number of hydrogen bond donors ($n_{\text{OHNH}} \leq 5$), and number of rotatable bonds ($n_{\text{rotb}} \leq 10$). Molecules **A**, **B** and **C** were constructed in the Molinspiration Online Property Calculation Software Toolkit (<http://www.molinspiration.com/>) for the evaluation of the Lipinski's Rule. The results are shown in Table 6 and proposed compound did not violate the Lipinski's Rule of Five.

Table 6. Calculated parameters of the Lipinski's Rule for the proposed molecules.

Compound	miLogP	MW	n _{ON}	n _{OHNH}	n _{rotb}	n _{violations}
A	2.38	420.47	8	1	6	0
B	4.01	528.40	8	1	7	1
C	3.26	479.53	9	1	8	0

Due to the fact that the proposed compounds presented predicted activity and docking interaction energy values higher than the values of the compounds of the series evaluated, these new molecules can be considered propitious for the performance as drugs. In addition, these compounds are unlikely to present bioavailability problems.

Conclusions

A series of 77 *Leishmania donovani* inhibitors was selected from the literature for a 4D-QSAR study, applying three test groups and seven alignments. The best model, Model A1, presents important features that can be applied in the development of new NMT inhibitors. The Model A1 showed r^2 , q^2 , and values of 0.76 and 0.70 respectively. In addition, it has external validation values of $R^2_{\text{pred}} = 0.74$ and $R^2_{\text{m (test)}} = 0.54$. Moreover, based on the descriptors obtained from Model A1, Compounds **A**, **B** and **C** were proposed, which demonstrated biological activity values superior to those of the most active compound. These results were corroborated by docking studies and the ADME evaluation. Consequently, the proposed compounds may be considered as promising drug candidates for the treatment of leishmaniasis.

Acknowledgements

We thank the Brazilian agencies “Conselho Nacional de Desenvolvimento Científico e Tecnológico” (CNPq, Brazil), “Coordenação de Aperfeiçoamento de Pessoal de Nível

Superior” (CAPES) and “Fundação de Amparo à Pesquisa do Estado de Minas Gerais” (FAPEMIG, Brazil) for their support. We thank A. J. Hopfinger who kindly supplied the 4D-QSAR program for academic use.

References

1. WHO - World Health Organization; *WHO and Gilead Sciences extend collaboration against visceral leishmaniasis*. Geneva, 2016.
2. de Menezes, J. P. B.; Guedes, C. E. S.; Petersen, A.; Fraga, D. B. M.; Veras, P. S. T.; *Biomed Research International*, **2015**, *2015*, 815023.
3. Ghosh, S.; Kar, N.; Bera, T.; *International Journal of Biological Macromolecules* **2016**, *93*, 961.
4. Goldston, A. M.; Sharma, A. I.; Paul, K. S.; Engman, D. M.; *Trends in Parasitology* **2014**, *30*, 350.
5. Brannigan, J. A.; Smith, B. A.; Yu, Z. Y.; Brzozowski, A. M.; Hodgkinson, M. R.; Maroof, A.; Price, H. P.; Meier, F.; Leatherbarrow, R. J.; Tate, E. W.; Smith, D. F.; Wilkinson, A. J.; *Journal of Molecular Biology* **2010**, *396*, 985.
6. Leatherbarrow, R.; Tate, E.; Yu, Z.; Rackham, M.; *WO pat. 2013/083991*, **2013**.
7. Release; *HyperChem Release 7; HyperCube Inc*; Gainesville, Florida, 2002.
8. Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P.; *Journal of Computational Chemistry* **2006**, *27*, 1101.
9. Hopfinger, A.; *4D-QSAR Package User's Manual*; In *The Chem21 Group, Inc.:1780 Wilson Drive*; Lake Forest, USA, 2001.
10. Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C.; *Journal of the American Chemical Society* **1997**, *119*, 10509.
11. Romeiro, N. C.; Albuquerque, M. G.; de Alencastro, R. B.; Ravi, M.; Hopfinger, A. J.; *Journal of Computer-Aided Molecular Design* **2005**, *19*, 385.
12. Rogers, D.; Hopfinger, A. J.; *Journal of Chemical Information and Computer Sciences* **1994**, *34*, 854.
13. Allinger, N. L.; *Journal of the American Chemical Society* **1977**, *99*, 8127.
14. Kiralj, R.; Ferreira, M. M. C.; *Journal of the Brazilian Chemical Society* **2009**, *20*, 770.
15. Kar, S.; Roy, K.; *Indian Journal of Biochemistry & Biophysics* **2011**, *48*, 111.
16. Roy, K.; Paul, S.; *Qsar & Combinatorial Science* **2009**, *28*, 406.
17. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N.; *Journal of Computational Chemistry* **2013**, *34*, 1071.
18. Veerasamy, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., Agrawal, R. K.; *International Journal of Drug Design & Discovery*, **2011**, *3*, 511.
19. Thomsen, R.; Christensen, M. H.; *Journal of Medicinal Chemistry* **2006**, *49*, 3315.
20. Sodero, A. C. R.; Romeiro, N. C.; da Cunha, E. F. F.; Magalhaes, U. D.; de Alencastro, R. B.; Rodrigues, C. R.; Cabral, L. M.; Castro, H. C.; Albuquerque, M. G.; *Molecules* **2012**, *17*, 7415.
21. Patel, H. M.; Sing, B.; Bhardwaj, V.; Palkar, M.; Shaikh, M. S.; Rane, R.; Alwan, W. S.; Gadad, A. K.; Noolvi, M. N.; Karpoomath, R.; *European Journal of Medicinal Chemistry* **2015**, *93*, 599.

Supporting Information Available

Table S1. Descriptors generated by Model A1.

	0,-2,4,np	0,3,2,any	3,-2,-1,np	0,-3,2,any	0,4,-1,any
1	0.000	1.002	0.085	0.767	0.001
2	0.736	0.964	0.085	0.000	0.012
3	0.004	0.968	0.000	0.000	0.002
4	0.000	0.000	0.000	0.000	0.104
5	0.001	0.966	0.098	0.000	0.003
6	0.654	0.986	0.000	0.000	1.002
7	0.576	0.978	0.000	0.000	0.827
8	0.516	0.943	0.010	0.000	0.006
9	0.555	0.956	0.000	0.000	0.001
10	0.671	0.966	0.000	0.000	0.006
11	0.749	0.962	0.000	0.000	0.003
12	0.398	0.934	0.000	0.000	0.009
13	0.680	0.956	0.000	0.000	0.000
14	0.734	0.965	0.000	0.000	0.000
15	0.764	0.960	0.000	0.000	0.000
16	0.005	0.963	0.000	0.000	0.000
17	0.001	0.973	0.000	0.000	0.005
18	0.491	0.967	0.000	0.000	0.005
19	0.469	0.979	0.000	0.000	0.006
20	0.361	0.972	0.000	0.000	0.003
21	0.641	0.971	0.000	0.000	0.008
22	0.393	0.956	0.000	0.000	0.005
23	0.672	0.977	0.000	0.000	0.003
24	0.638	0.972	0.000	0.000	0.005
25	0.618	0.959	0.000	0.000	0.003
26	0.203	0.983	0.066	0.644	0.006
27	0.210	0.972	0.000	0.646	0.003
28	0.241	0.981	0.000	0.644	0.004
29	0.179	0.981	0.030	0.640	0.005
30	0.158	0.964	0.000	0.563	0.000
31	0.183	0.965	0.000	0.566	0.000
32	0.162	0.964	0.000	0.531	0.000
33	0.163	0.967	0.000	0.538	0.000
34	0.178	0.964	0.000	0.574	0.000
35	0.269	0.968	0.000	0.637	0.007
36	0.220	0.976	0.044	0.632	0.005
37	0.193	0.967	0.000	0.640	0.004
38	0.244	0.981	0.000	0.671	0.004
39	0.171	0.984	0.071	0.655	0.003
40	0.254	0.969	0.000	0.646	0.008
41	0.234	0.964	0.000	0.655	0.007

42	0.000	0.976	0.000	0.203	0.978
43	0.000	0.992	0.040	0.244	0.975
44	0.000	0.987	0.005	0.204	0.975
45	0.000	0.980	0.000	0.219	0.857
46	0.000	0.984	0.000	0.195	0.827
47	0.000	0.990	0.000	0.205	0.809
48	0.000	0.987	0.000	0.201	0.929
49	0.000	0.988	0.000	0.176	0.912
50	0.000	0.977	0.000	0.172	0.002
51	0.000	0.965	0.002	0.227	0.001
52	0.000	0.979	0.000	0.169	0.003
53	0.000	0.029	0.000	0.205	0.001
54	0.000	0.016	0.070	0.254	0.009
55	0.000	1.000	0.000	0.187	0.774
56	0.000	0.993	0.000	0.000	0.730
57	0.000	0.999	0.000	0.000	0.745
58	0.000	0.998	0.000	0.000	0.742
59	0.000	0.997	0.465	0.119	0.003
60	0.000	0.049	0.683	0.266	0.003
61	0.000	0.031	0.690	0.230	0.004
62	0.000	0.116	0.544	0.226	0.003
63	0.000	0.099	0.654	0.239	0.006
64	0.000	0.965	0.426	0.186	0.000
65	0.000	0.965	0.650	0.179	0.000
66	0.000	0.993	0.576	0.228	0.723
67	0.000	0.993	0.603	0.247	0.917
68	0.000	0.999	0.360	0.222	0.901
69	0.000	0.988	0.395	0.196	0.006
70	0.000	0.026	0.636	0.263	0.002
71	0.000	1.005	0.696	0.198	0.840
72	0.000	0.116	0.635	0.223	0.003
73	0.000	0.987	0.707	0.270	0.010
74	0.000	0.998	0.108	0.151	0.776
75	0.000	1.000	0.648	0.780	0.000
76	0.000	1.002	0.528	0.787	0.000
77	0.000	1.001	0.034	0.789	0.001

Table S2 Experimental and predicted pIC₅₀ values and residuals values (pIC₅₀predicted - pIC₅₀experimental) for compounds of Model A1.

Compound	pIC₅₀ Experimental	pIC₅₀ Predicted	Residuals
1	6.301	6.887	0.586
2	6.222	6.215	-0.008
3	5.456	5.227	-0.229
4	6.097	5.666	-0.431
5	5.377	5.367	-0.010
6	5.031	5.044	0.013
7	4.633	5.121	0.488
8	5.366	5.861	0.495
9	5.796	5.890	0.094
10	5.155	6.018	0.863
11	5.155	6.116	0.961
12	6.000	5.708	-0.292
13	6.678	6.040	-0.638
14	6.769	6.099	-0.670
15	6.481	6.137	-0.344
16	5.796	5.233	-0.563
17	5.523	5.218	-0.305
18	6.509	5.804	-0.705
19	6.155	5.770	-0.385
20	5.796	5.649	-0.148
21	5.620	5.977	0.357
22	5.398	5.694	0.296
23	6.155	6.016	-0.139
24	5.824	5.976	0.152
25	6.301	5.962	-0.340
26	7.180	6.856	-0.324
27	7.356	6.782	-0.575
28	6.854	6.808	-0.046
29	6.097	6.769	0.672
30	7.155	6.558	-0.597
31	6.824	6.594	-0.231
32	6.523	6.498	-0.025
33	6.000	6.512	0.512
34	5.682	6.604	0.922
35	7.337	6.832	-0.505
36	7.509	6.825	-0.684
37	7.137	6.751	-0.386
38	6.409	6.867	0.458
39	7.081	6.850	-0.231
40	6.620	6.831	0.211
41	6.553	6.829	0.276
42	4.468	4.707	0.239
43	3.870	4.842	0.972
44	5.236	4.713	-0.523
45	4.521	4.852	0.331

46	4.292	4.829	0.537
47	5.468	4.863	-0.605
48	5.276	4.743	-0.533
49	4.692	4.708	0.016
50	4.769	5.567	0.798
51	5.347	5.689	0.342
52	5.113	5.559	0.446
53	6.060	6.164	0.104
54	6.337	6.365	0.028
55	4.418	4.854	0.436
56	4.252	4.520	0.268
57	4.661	4.502	-0.159
58	4.661	4.505	-0.156
59	5.796	6.125	0.329
60	6.569	7.271	0.702
61	7.699	7.217	-0.482
62	7.070	6.949	-0.121
63	7.301	7.143	-0.158
64	6.409	6.225	-0.184
65	7.000	6.538	-0.462
66	6.769	5.830	-0.939
67	5.319	5.724	0.405
68	6.167	5.331	-0.836
69	5.677	6.182	0.505
70	7.004	7.210	0.206
71	5.509	5.826	0.317
72	6.886	7.076	0.190
73	6.187	6.784	0.597
74	5.347	4.938	-0.410
75	7.757	7.737	-0.021
76	8.000	7.575	-0.426
77	7.155	6.858	-0.298
