

UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS BIOLÓGICAS E AMBIENTAIS
GRADUAÇÃO EM BIOTECNOLOGIA

VITOR MORAES MACHADO DE OLIVEIRA

**MODELO PREDITIVO QSAR PARA AVALIAÇÃO DA ATIVIDADE DE
INIBIDORES DE ACETILCOLINESTERASE**

DOURADOS-MS

2020

VITOR MORAES MACHADO DE OLIVEIRA

**MODELO PREDITIVO QSAR PARA AVALIAÇÃO DA ATIVIDADE DE
INIBIDORES DE ACETILCOLINESTERASE**

Trabalho de Conclusão de Curso apresentado à
Universidade Federal da Grande Dourados para
obtenção do título de Bacharel em Biotecnologia, sob
orientação do Prof. Dr. Heberth Juliano Vieira.

DOURADOS-MS

2020

VITOR MORAES MACHADO DE OLIVEIRA

**MODELO PREDITIVO QSAR PARA AVALIAÇÃO DA ATIVIDADE DE
INIBIDORES DE ACETILCOLINESTERASE**

Trabalho de Conclusão de Curso aprovado pela Banca Examinadora como requisito parcial para obtenção do título de Bacharel em Biotecnologia, da Universidade Federal da Grande Dourados.

Orientador: Dr. Heberth Juliano Vieira

Área de Concentração: ...

Aprovado em: *(colocar data da defesa)*

BANCA EXAMINADORA

Prof. Dr. Heberth Juliano Vieira
Presidente

Prof.^a Dr.^a Daniele Mignolo
Membro

Prof. Dr. Thiago Sequinel
Membro

AGRADECIMENTOS

Este trabalho é fruto de uma esplêndida cooperação entre aluno e professor, orientando e orientador, por isso devo agradecimentos ao meu orientador Heberth que me ajudou com sua sabedoria ímpar, orientou em todos os detalhes técnicos do trabalho e principalmente me motivou a dedicar-me cada vez mais para a conclusão deste estudo.

Nada seria possível sem o apoio familiar que tive, de meus pais Tito e Maria e meus irmãos Vinicius e Inara que nunca deixaram de me apoiar, e que principalmente sempre confiaram nas minhas competências, por isso os agradeço profundamente.

Minha namorada Caroline foi o maior e mais incrível presente que recebi durante o período em que vivi em Dourados, com ela vivi momentos fantásticos e inesquecíveis, e foi quem mais me ajudou de perto e auxiliou em todas as dificuldades que passei durante a vida acadêmica, obrigado por tudo, meu amor.

Não poderia deixar de agradecer aos meus melhores amigos, que me proporcionaram os momentos inigualáveis, de muita diversão e loucura, experiências que tornaram a cidade de Dourados onde mais eu gostaria de estar. Meus agradecimentos eternos a Lucas Leite, Henrique, Mateus Barbosa, Rodrigo Ribeiro, Anderson Hajime, Mateus Ferreira, Héverton, Rafael Santelli, Rafael Yudi, Guilherme, Kenji, Matheus Pereira, Murilo Stein, Amanda, José Edson, André, Gabriel, Bethânia, Fernanda e Joaquim.

Gostaria por fim de agradecer a banca avaliadora que se disponibilizou a analisar meu trabalho de conclusão de curso e as dependências da UFGD que me proporcionaram a melhor estrutura para o meu desenvolvimento pessoal e acadêmico.

RESUMO

A doença de Alzheimer é um distúrbio neurodegenerativo progressivo associado à perda progressiva de memória que aflige milhões de pessoas no mundo, e no momento apenas uma classe de medicamentos é utilizada para o tratamento dos sintomas de DA, os inibidores de colinesterase (ChE). Neste trabalho avaliou-se os dados da atividade biológica (IC50) de potenciais inibidores de acetilcolinesterase (AChE) sendo 07 derivados de Piperidinona, 7 derivados de ácido catívico e um derivado de tacrina, apresentados em PARLAR (2019), ALZA et al. (2018) e BITAM et al (2018), respectivamente, para a obtenção de um modelo QSAR preditivo empregando calibração MRL. Os descritores moleculares dos derivados sintetizados pelos autores foram calculados empregando o software PADEL. A seleção dos descritores com maior influência na atividade biológica (IC50) das substâncias foi calculada empregando Microsoft Excel®. A obtenção do modelo preditivo com os descritores foi realizada empregando o programa BuidQSAR. Os descritores calculados englobam propriedades moleculares com aspectos topológicos, eletrônicos e geométricos. A importância de cada descritor foi atribuída de acordo com sua influência na atividade biológica da molécula. Depois de obtida a lista de descritores moleculares, verificou-se quais desses descritores apresentavam maiores correlação com a atividade biológica das substâncias, aqueles com correlação absoluta superior a 0,1805 foram selecionados para a construção do modelo preditivo e transferidos ao BuildQSAR. Para a calibração multivariada utilizando MRL, o melhor modelo QSAR mostrou-se estável, robusto, com valores residuais baixos e preditivo para a atividade biológica com base nas substâncias estudadas, com valores de $q^2= 0,986$, $r^2=0,996$ e $r^2(\text{ajustado})= 0,992$ associados aos descritores MATS6v, XlogP, ATSC8c e ZMIC1. A equação preditiva apresentou os seguintes coeficientes: $pIC_{50} = - 1,0107 (\pm 0,0893)$ MATS6v - 0,3741 ($\pm 0,0918$) XlogP + 0,1478 ($\pm 0,0853$) ATSC8c - 0,0941 ($\pm 0,0878$) ZMIC1 - 0,7575 ($\pm 0,0744$) Todos os coeficientes apresentaram significância estatística ao nível de 95% de confiança. Os coeficientes de correlação entre os descritores obtidos foram inferiores

a 0,450. Os resultados apresentados demonstraram que o modelo preditivo obtido permitiu a obtenção de modelos confiáveis e preditivos, fornecendo informações para a síntese de novas substâncias baseadas na estrutura e potencialmente mais ativas na inibição de acetilcolinesterase.

ABSTRACT

Alzheimer's disease is a progressive neurodegenerative disorder associated with progressive memory loss that afflicts millions of people worldwide, and at the moment only one class of drugs is used to treat symptoms of AD, cholinesterase inhibitors (ChE). In this work, the biological activity data (IC₅₀) of potential acetylcholinesterase inhibitors (AChE) were evaluated, being 07 derivatives of Piperidinone, 7 derivatives of cativic acid and one derived from tacrine, described in the literature. to obtain a predictive QSAR model employing multiple linear regression. The molecular descriptors of the derivatives synthesized by the authors were calculated using the PADEL software. The selection of descriptors with the greatest influence on the biological activity (IC₅₀) of the substances was calculated using Microsoft Excel®. The predictive model was obtained using the descriptors using the BuidQSAR program. The calculated descriptors include molecular properties with topological, electronic and geometric aspects. The importance of each descriptor was attributed according to its influence on the biological activity of the molecule. After obtaining the list of molecular descriptors, it was found which of these descriptors had the highest correlation with the biological activity of the substances, those with an absolute correlation greater than 0.1805 were selected to build the predictive model and transferred to BuildQSAR. For multivariate calibration using MRL, the best QSAR model proved to be stable, robust, with low residual values and predictive for biological activity based on the substances studied, with values of $q^2 = 0.986$, $r^2 = 0.996$ and r^2 (adjusted) = 0.992 associated with MATS6v, XlogP, ATSC8c and ZMIC1. The predictive equation presented the following coefficients: $pIC_{50} = - 1.0107 (\pm$

0.0893) MATS6v - 0.3741 (\pm 0.0918) XlogP + 0.1478 (\pm 0.0853) ATSC8c - 0.0941 (\pm 0 , 0878) ZMIC1 - 0.7575 (\pm 0.0744) All coefficients showed statistical significance at the 95% confidence level. The correlation coefficients between the descriptors obtained were less than 0.450. The results presented showed that the predictive model obtained allowed to obtain reliable and predictive models, providing information for the synthesis of new substances based on the structure and potentially more active in the inhibition of acetylcholinesterase.

INTRODUÇÃO

O método preditivo de atividade biológicas ou propriedades físico-químicas a partir das relações quantitativas das estruturas químicas (QSAR/QSPR) é uma área que possui destaque na comunidade científica nos ramos da quimioinformática, bioinformática e quimiofarmacêutica teórica. Esta técnica utiliza de descritores moleculares baseados em propriedades estruturais, físico-químicas e conformacionais de potenciais agentes previamente testados para o desenvolvimento de uma equação matemática que revela a predição.

Os descritores moleculares são essenciais para o desenvolvimento de um modelo QSAR, são calculados a partir da estrutura molecular da substância. Os descritores molecular identificam as forças intermoleculares que governam as interações fármaco-receptor e podem ser distinguidos pela dimensionalidade da representação estrutural. Cada descritor molecular traduz uma propriedade físico-química da molécula, podendo ser classificados como: descritores constitucionais, descritores topológicos, índices de conectividade, índices E-state, descritores de autocorrelação, carga descritores, propriedades moleculares, índices de forma *kappa*, descritores de tipo MOE e impressões digitais (*fingerprints*).

Relações QSAR são úteis para compreender e definir mecanismos de ação de moléculas, investigar possíveis interações entre sistemas, revelando a viabilidade de realizar-

se um determinado experimento, que por sua vez envolve comitês de ética, prática com animais e é financeiramente mais custoso (SILVA, 2017). O sucesso da aplicação de qualquer modelo QSAR necessita da precisão dos dados de entrada, da seleção de descritores apropriados, ferramentas estatísticas e, principalmente, da ratificação do modelo desenvolvido (BITAM et al, 2018).

Dados do relatório da ADI (do inglês, Alzheimer's Disease International) de 2013 apontam que nos próximos 40 anos o número de idosos que necessitarão de cuidados deve triplicar, alcançando o número de aproximadamente 277 milhões até 2050, aumentando expressivamente o número de pessoas com a doença de Alzheimer (DA). A doença de Alzheimer é um distúrbio neurodegenerativo progressivo associado à perda progressiva de memória, um declínio nas habilidades de linguagem e outras deficiências cognitivas (ALZA et al, 2014). E no momento apenas uma classe de medicamento é utilizada para o tratamento dos sintomas de DA, são estes os inibidores de acetilcolinesterase (AChE) e butirilcolinesterase (BuChE) (BITAM et al, 2018).

As colinesterases (ChE) são enzimas encarregadas pela degradação da acetilcolina, neurotransmissor responsável pelo controle de impulsos nervosos destinados aos músculos do organismo. Em um indivíduo portador da doença de Alzheimer a quantidade de acetilcolina em nível cerebral será reduzida, tanto por falta de produção, tanto por demasiada degradação realizada por colinesterases, por isso o tratamento com anticolinesterásicos é o mais visado para modelos preditivos. Atualmente os medicamentos mais utilizados estão descritos na tabela 1 abaixo.

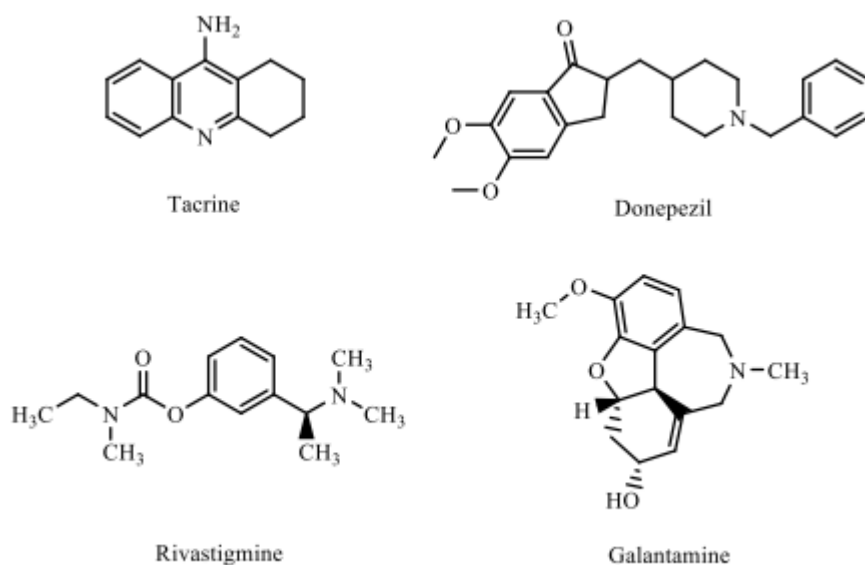


Imagem 1: Tacrina, donepezila, rivastigmina e galantamina. Compostos aprovados para o tratamento da DA pela FDA (PARLAR, 2019)

As moléculas escolhidas para o estudo de modelo QSAR são potenciais inibidores de ChE, sendo elas derivados de ácido catívico, que são encontrados com facilidade em plantas do gênero *Grindelia*, Sabe-se que as plantas desse gênero são fontes de compostos bioativos, principalmente diterpenóides do tipo labdano e derivados do óxido de mano, mono- e sesquiterpenos, poliacetilenos, flavonóides e saponinas. As atividades anti-inflamatórias, expectorantes, antiespasmódicas e antimicrobianas, bem como atividade anti-alimentar para os insetos, foram relatados para extratos ou metabólitos secundários obtidos de plantas de *Grindelia* (ALZA et al, 2014), e derivados de piperidinona, compostos que são encontrados em muitas plantas medicinais, ao redor de todo o globo, como *Conium maculatum*. (PARLAR, 2019).

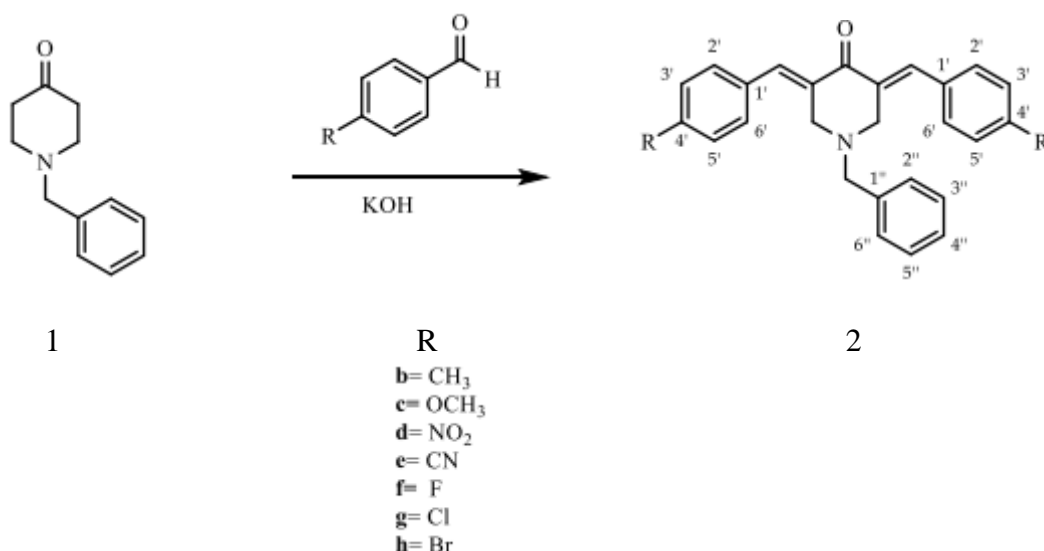
OBJETIVO

O objetivo deste trabalho foi construir um modelo matemático preditivo, utilizando descritores moleculares previamente selecionados, para a avaliação *in silico* da atividade biológica de compostos com potencial de inibição de Acetilcolinesterase.

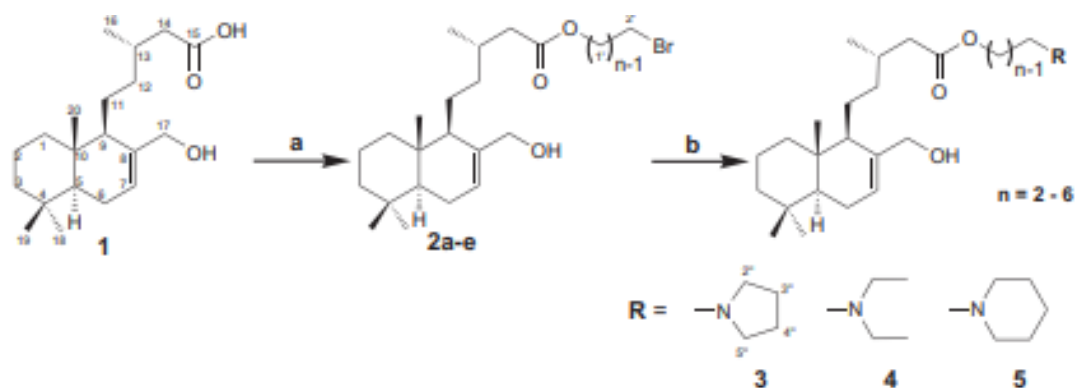
METODOLOGIA

Seleção das moléculas.

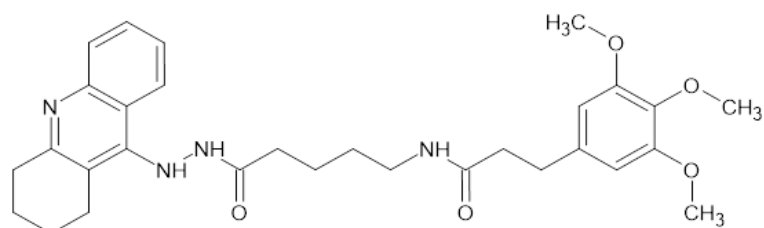
As moléculas deste estudo foram selecionadas após pesquisa na literatura, e os estudos de **PARLAR de 2019 e ALZA de 2018** demonstraram que os derivados de piperidinona e ácido catívico, respectivamente, são potenciais inibidores de AChE pouco explorados até o momento. De cada artigo foram selecionadas 7 moléculas, sendo elas com potencial inibitório 50% (IC50) mais elevado, como descrevem os esquemas 1 e 2. Os potenciais inibitórios destas moléculas foram emparelhados ao potencial de um derivado de tacrina descrito em BITAM et al, 2018, como demonstra o esquema 3. As informações das substâncias estão disponibilizadas no ANEXO.



Esquema 1: fórmulas estruturais dos derivados de piperidinona segundo PARLAR, 2019



Esquema 2: fórmulas estruturais dos derivados de ácido catívico segundo ALZA et al, 2018



Esquema 3: fórmula estrutural do derivado de tacrina segundo BITAM et al, 2018

Obtenção das estruturas.

As moléculas foram obtidas seguindo os esquemas descritos acima, utilizando o software ACD/ChemSketch versão freeware 2019 2.1. Os arquivos foram salvos no formato .mdl (.mol) para futura obtenção dos descritores moleculares.

Obtenção dos descritores moleculares.

As moléculas obtidas no formato .mol no foram submetidas ao software Padel-Descriptor versão 2.2 para o cálculo dos descritores 1D e 2D. Os descritores 3D e fingerprints não foram utilizados neste trabalho. Todos os 1444 descritores de cada molécula foram salvos em um único arquivo .csv. A tabela completa com todos os descritores moleculares está no ANEXO II.

Pré-seleção dos descritores moleculares.

O programa Excel versão 1808 (Office 2019) foi empregado para executar o arquivo .csv, incluir os valores de pIC50 de cada molécula e realizar uma pré-seleção dos descritores moleculares mais interessantes para realização do modelo QSAR. Os descritores moleculares calculados na qual a média e desvio padrão entre todas as moléculas são iguais a zero foram removidos, assim como os descritores que não possuíam correlação absoluta mínima com o valor de pIC50 descrita pela fórmula: $\frac{1}{15} \times \ln 15 = 0,1805$, logo, todos os descritores que possuíam um valor absoluto de correlação inferior a 0,1805 também foram removidos. (TARKO, 2017)

QSAR

O software escolhido para a obtenção e validação do modelo QSAR foi o BuidQSAR versão 2.1.0.0. Nesta etapa utilizou-se os descritores moleculares pré-selecionados anteriormente. Inicialmente, foi realizado um escalamento de dados, atualizando todos os descritores para uma mesma escala. Em seguida, variáveis foram selecionadas por algoritmo genético, sob os parâmetros de: 200 o número de gerações e 10 modelos por geração, selecionando 4 descritores para o modelo. E por fim foi realizada a regressão múltipla linear (MLR, do inglês Multiple Linear Regression) que exibiu os resultados estatísticos, referentes ao modelo, de equação preditiva, análise de variância, matriz de correlação e análise de coeficiente.

RESULTADOS E DISCUSSÃO

Seleção dos descritores moleculares

O software Padel-descriptor renderizou ao todo 1444 descritores moleculares, porém nem todos os descritores são relacionados ao modelo em estudo, e a presença destes pode provocar erros na seleção das variáveis para a construção do modelo, por isso seleção manual se fez necessária. Após a realização deste processo, obteve-se um total de 390 descritores

moleculares de interesse, e então tornou-se possível observar, algumas particularidades presentes nas estruturas das moléculas que são determinantes para inibição enzimática. Com isso, observou-se que a ausência de grupos ácidos, e presença de heteroátomos nos anéis ligados do derivado de tacrina, como foi descrito pelos descritores nBase e nF10Heteroring (TODESCHINI) certamente contribuem para que este composto tenha uma concentração inibitória mínima de 50% de Acetilcolinesterase inferior às demais, por exemplo.

Tabela 1: descrição segundo TODESCHINI em 2009 dos 20 descritores moleculares que possuem maior correlação com pIC50

Descritor	Descrição
nBase	Número de grupos de ácidos
MDEN-22	Distância entre nitrogênios de borda
nFG12Ring	Número de anéis ligados com mais de 12 membros
nTG12Ring	Número total de anéis com mais de 12 membros
nF10HeteroRing	Número de anéis ligados com 10 membros que contém heteroátomos
nFG12HeteroRing	Número de anéis ligados com mais de 12 membros que contém heteroátomos
nT10HeteroRing	Número total de anéis com 10 membros que contém heteroátomos
nTG12HeteroRing	Número total de anéis com mais de 12 membros que contém heteroátomos
EE_D	Índice tipo estrada (função log) da matriz de distância topológica
AATSC6v	Autocorrelação Broto-Moreau centrada média - lag 6 / ponderada pelos volumes de van der Waals
MATS6v	Autocorrelação de Moran de lag 6 ponderada pelo volume de van der Waals
MATS6p	Autocorrelação de Moran de lag 6 ponderada pela polarizabilidade
ATSC7e	Autocorrelação de Broto-Moreau centrada de lag 7 ponderada pela eletronegatividade de Sanderson
AATSC6p	Autocorrelação Broto-Moreau centrada média - lag 6 / ponderada por polarizabilidades
GATS3v	Autocorrelação Geary de lag 3 ponderada pelo volume de van der Waals
AATSC6m	Autocorrelação média de Broto-Moreau centrada - lag 6 / ponderada pela massa
AATSC6c	Autocorrelação média de Broto-Moreau centrada - lag 6 / ponderada pela carga
AATSC7e	Autocorrelação média de Broto-Moreau centrada - lag 7 / ponderada pela eletronegatividade de sanderson
ATSC6c	Autocorrelação de Broto-Moreau centrada - lag 6 / ponderada por cargas
GATS6v	Autocorrelação Geary de lag 6 ponderada pelo volume de van der Waals

Seleção das variáveis descritivas

Os descritores moleculares utilizados para a obtenção do modelo preditivo da atividade biológica pIC50 das substâncias inibidoras foram obtidos a partir de algoritmo genético e matriz de correlação, tecnologia esta que revelou várias combinações contendo 4 variáveis cada. A combinação MATS6v, XlogP, ATSC8c e ZMIC1, foi apontada pelo programa como a que possuía valor de R² e Q² mais próximos de 1, R² é o coeficiente de regressão e Q² o cálculo estatístico de relevância preditiva, o valor máximo para cada é 1, os valores mínimos toleráveis são 0,6 para R² e 0,5 para Q², uma diferença considerável de mais de 0,2 entre estes 2 valores podem significar que o modelo sofre de Overfitting, que significa que os valores se ajustam aos dados, porém não possui boa previsibilidade. Os valores foram respectivamente 0,996 e 0,986, e por isso o modelo foi construído a partir desta combinação.

Tabela 2: valores de pIC50 e dos descritores selecionados e escalonados para cada molécula.

Compostos	pIC50 ^a	MATS6v	XlogP	ATSC8c	ZMIC1
Ac catívico 3c	-0.507000000	0.004075064	-0.358727716	-0.620485821	-1.070995890
Ac catívico 3d	-0.744000000	0.145687038	-0.076434657	-0.323869660	-0.585662853
Ac catívico 3e	-1.290000000	0.080805810	0.205858402	-0.871896083	-0.088598948
Ac catívico 4b	-1.040000000	0.494050197	-0.372921222	-0.319499472	-1.682287776
Ac catívico 4c	-0.976000000	0.160880460	-0.090628162	-0.691138715	-1.268589853
Ac catívico 5b	-0.792000000	0.306772940	-0.358727716	-0.533019764	-1.070995899
Ac catívico 5e	-1.161000000	0.044000660	0.488151461	-0.879456500	0.418386906
Piperidinona b	-1.265000000	-0.351826990	2.472088157	0.284538564	0.522867935
Piperidinona c	-1.211000000	0.916333007	-0.401308233	2.083918270	0.567333846
Piperidinona d	-1.098000000	0.450356364	-0.860231587	-0.614698513	1.286903680
Piperidinona e	-1.300000000	0.379088734	-0.063818207	-1.423463905	0.439592873
Piperidinona f	-1.212000000	0.347588737	0.234245413	1.177909499	0.380278528
Piperidinona g	-1.256000000	0.280732118	0.450302112	0.739298260	0.818682803
Piperidinona h	-1.360000000	0.212948639	0.983347107	0.567550235	1.805092423
Tacrina	3.850000000	-3.471492787	-2.251195152	1.424313606	-0.472007767

^apIC50=-logIC50. Exposição dos descritores selecionados: XlogP corresponde ao coeficiente de partição de água de octanol de wang, ATSC8c corresponde à autocorrelação de lag 8

ponderada pela carga de gasteiger e ZMIC1 ao Índice de conteúdo de informação modificado por Z (simetria de vizinhança de 1 ordem).

Matriz de correlação dos coeficientes

Nos estudos QSAR a matriz de correlação traduz uma correspondência entre os valores dos descritores moleculares e a atividade biológica dos compostos. valores muito altos de R indicam uma alta proximidade, sendo considerados repetitivos por serem extremamente correlacionados, portanto altos valores de R são indesejáveis. No entanto, a faixa de valores a serem considerados deve ser definida de acordo com o estudo a ser realizado (SANTOS, 2017). Em estudos de QSAR recomenda-se valores de correlação dos descritores com a atividade biológica entre 0,3 e 0,6 (FERREIRA et al, 2002)

Tabela 3: matriz de correlação das variáveis

	MATS6v	XlogP	ATSC8c	ZMIC1
MATS6v	1	0.450	0.272	0.145
XlogP	0.450	1	0.110	0.355
ATSC8c	0.272	0.110	1	0.276
ZMIC1	0.145	0.355	0.276	1

Com base nos valores apresentados na tabela 3, e nos valores de R^2 e Q^2 , pode-se afirmar que o algoritmo genético foi preciso na determinação das variáveis, que por mais que apresentem correlação elevada com os valores de pIC50, entre os descritores o valor de correlação se manteve abaixo de 0,6, portanto, sem correlação excessiva.

Equação preditiva

A regressão múltipla linear computou os cálculos estatísticos referentes ao modelo e está apresentado no esquema 4. A equação preditiva é descrita como $pIC50 = \beta_1 \times desc\ 1 + \beta_2 \times desc\ 2 + \beta_3 \times desc\ 3 + \beta_4 \times desc\ 4 + \beta_0$; onde β_n são os coeficientes e “descr n” são as variáveis descritivas do modelo matemático.

$$pIC50 = - 1,0107 (\pm 0,0893) MATS6v - 0,3741 (\pm 0,0918) XlogP + 0,1478 (\pm 0,0853) ATSC8c - 0,0941 (\pm 0,0878) ZMIC1 - 0,7575 (\pm 0,0744)$$

$$(n = 15 ; R = 0,996 ; s = 0,129 ; F = 349,344 ; p < 0,0001 ; Q^2 = 0,986 ; SPress = 0,184 ; SDEP = 0,155)$$

Esquema 4: Equação preditiva e parâmetros estatísticos

Esta equação primordialmente traduz a devida correlação entre a atividade inibitória pIC50 e os descritores moleculares selecionados. Pode-se observar que os descritores MATS6v, XlogP e ZMIC1 são inversamente proporcionais ao valor de pIC50, ou seja, quanto maior o valor desses descritores, menor será o valor de pIC50, enquanto ATSC8c é diretamente proporcional, logo, quanto maior seu valor, menor é a concentração inibitória de 50% da enzima Acetilcolinesterase (AChE).

A equação preditiva também traz a informação do quão preditivo é o modelo, pois a partir dos descritores selecionados, calcula-se o pIC50, e, quanto mais os resultados da equação se aproximarem dos resultados experimentais, maior é a capacidade preditiva do modelo. Os resultados da equação estão presentes na tabela de resíduo abaixo.

Compostos	pIC50(observado)	pIC50(calculado)	pIC50(resíduo)	Desvio (res)
Ac catívico 3c	-0,507	-0,618	0,111	0,860
Ac catívico 3d	-0,744	-0,869	0,125	0,965
Ac catívico 3e	-1,290	-1,037	-0,253	-1,958
Ac catívico 4b	-1,040	-1,006	-0,034	-0,262
Ac catívico 4c	-0,976	-0,869	-0,107	-0,828

Ac catívico 5b	-0,792	-0,911	0,119	0,922
Ac catívico 5e	-1,161	-1,154	-0,007	-0,054
Piperidinona b	-1,265	-1,334	0,069	0,533
Piperidinona c	-1,211	-1,279	0,068	0,524
Piperidinona d	-1,098	-1,103	0,005	0,037
Piperidinona e	-1,300	-1,369	0,069	0,530
Piperidinona f	-1,212	-1,058	-0,154	-1,189
Piperidinona g	-1,256	-1,177	-0,079	-0,607
Piperidinona h	-1,360	-1,427	0,067	0,515
Tacrina	3,850	3,848	0,002	0,012

Tabela 4: tabela de resíduo

A tabela de resíduo, compara os valores de pIC50 experimental, e o calculado por meio da equação preditiva QSAR; o resultado desta comparação é o resíduo, presente na 4ª coluna, que é a subtração entre estes valores. Os resultados do resíduo revelam que esta variação é inferior a 22% para os derivados de Ác catívico, inferior a 12% para os derivados de piperidinona, e de 0,0005% para o derivado da tacrina.

O gráfico 1 apresenta os dados da tabela residual, onde os valores de pIC50 observados experimentalmente e aqueles calculados pela equação preditiva obtida neste trabalho são plotados para análise visual da compatibilidade dos resultados. Ao analisar o gráfico e a tabela de resíduo, percebe-se que a regressão linear é alta, e baixa a variação entre os valores observados e calculados, o que indica que a equação preditiva é adequada ao modelo QSAR para inibição de AChE.

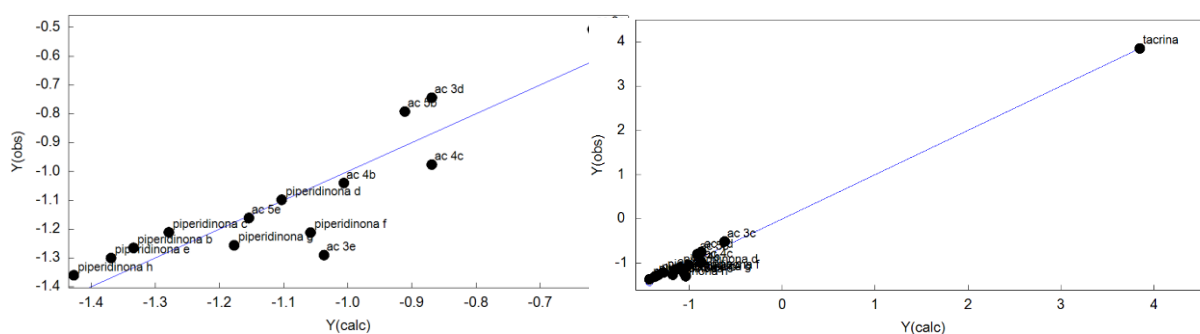


Gráfico 1: gráfico de relação y observado / y calculado

Análise dos coeficientes

A equação preditiva permite o cálculo do valor de pIC50 a partir dos descritores, porém a análise visual é dificultada pela presença de mais de uma variável descritiva. Neste sentido, para avaliar a significância dos coeficientes, é realizado um cálculo estatístico para comprovar a linearidade dos dados e a precisão da equação, representado pela tabela abaixo.

Tabela 5: Análise dos coeficientes

Preditor	coeficiente	Stdev	95% conf.	t-ratio	p
Constante	-0,7575	0,0334	0,0744	-22,6746	0,0000
MATS6v	-1,0107	0,0401	0,0893	-25,2032	0,0000
XlogP	-0,3741	0,0412	0,0918	-9,0775	0,0000
ATSC8c	0,1478	0,0383	0,0853	3,8595	0,0027
ZMIC1	-0,0941	0,0394	0,0878	-2,3879	0,0360

Pode-se observar que os valores de P, para o coeficiente de cada descritor molecular é próximo a 0, e inferior a 0,05, o que indica que os valores são significativos com 95% de confiabilidade, comprovando-se assim, a precisão dos coeficientes.

Análise de variância da equação de regressão linear múltipla

Neste caso, a análise de variância (anova), indica se o valor de pIC50 é dependente dos valores dos descritores moleculares selecionados ou não, indicando a significância estatística dos descritores. O valor de F calculado, é de 349,34, o que é considerado elevado e superior ao

F tabelado, sendo significativa a diferença entre os descritores a 0,05 de confiabilidade, o que representa a dependência dos descritores empregados por parte de pIC50.

Tabela 6: Análise de variância

	G.L	SQ	QM	F	P
Regreção	4	23,3912	5,8478	349,3437	0,0000
Erro	10	0,1674	0,0167		
Total	14	23,5586			

CONCLUSÃO

Após todas as análises e procedimentos realizados, pode-se afirmar que o objetivo inicial foi alcançado: um modelo QSAR foi desenvolvido seguindo os procedimentos descritos. Os coeficientes foram significativos, os valores residuais de pIC50 calculados são diminutos e a correlação dos dados é expressiva, e dessa maneira conclui-se que os descritores moleculares e o modelo possuem alta capacidade preditiva.

O estudo principalmente indicou, as melhores características moleculares que uma molécula deve conter para uma maior efetividade na inibição de acetilcolinesterase, como o volume molecular(MATS6c), lipoficidade (XlogP), e ramificação molecular(ZMIC1), que devem ser reduzidos, enquanto a carga parcial deverá possuir valor mais elevado(ATSC8c). Este resultado, e produções de modelos QSAR futuros auxiliam na formulação de novas moléculas para produção sintética e futuros testes em laboratório, uma etapa fundamental na geração de novos medicamentos contra a doença de Alzheimer (DA).

REFERÊNCIAS

SILVA, Mônica de Abreu. Modelos preditivos baseados em descritores moleculares e modos de interação receptor-ligante para inibidores de Acetilcolinesterase. 2017.

BITAM, S.; HAMADACHE, M.; HANINI, S. Prediction of therapeutic potency of tacrine derivatives as BuChE inhibitors from quantitative structure–activity relationship modelling. *SAR and QSAR in Environmental Research*, v. 29, n. 3, p. 213-230, 2018.

ALZHEIMER'S DISEASE INTERNATIONAL. World Alzheimer Report 2013: Journey of Caring: An analysis of long-term care for dementia. Alzheimer's Disease International, 2013.

ALZA, Natalia P. et al. Synthesis and cholinesterase inhibition of cativic acid derivatives. *Bioorganic & Medicinal Chemistry*, v. 22, n. 15, p. 3838-3849, 2014.

KIA, Yalda et al. Synthesis and discovery of novel piperidone-grafted mono-and bis-spirooxindole-hexahydropyrrolizines as potent cholinesterase inhibitors. *Bioorganic & medicinal chemistry*, v. 21, n. 7, p. 1696-1707, 2013.

PARLAR, Sulunay. Synthesis and cholinesterase inhibitory activity studies of some piperidinone derivatives. **Organic Communications**, v. 12, n. 4, p. 209, 2019.

YAP, Chun Wei. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. **Journal of computational chemistry**, v. 32, n. 7, p. 1466-1474, 2011.

DE OLIVEIRA, Daniel Barbosa; GAUDIO, Anderson Coser. BuildQSAR: a new computer program for QSAR analysis. **Quantitative Structure-Activity Relationships: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis**, v. 19, n. 6, p. 599-601, 2000.

TARKO, Laszlo. A Selection Method for Molecular Descriptors and QSPR Equations. *MATCH Communications in Mathematical and in Computer Chemistry*, v. 77, p. 245-72, 2017.

TODESCHINI, Roberto; CONSONNI, Viviana. **Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references**. John Wiley & Sons, 2009.

SANTOS, Gênisson dos Reis. Um estudo de QSAR sobre a interação de compostos arilpiperazínicos com o receptor 5-HT_{2a} utilizando os métodos PLS e ANN. 2017. Tese de Doutorado. Universidade de São Paulo.

FERREIRA, M.; MONTANARI, C.; GAUDIO, A. Seleção de variáveis em QSAR. **Química Nova**, v. 25, n. 3, p. 439-448, 2002.

Tabela completa com os dados utilizados:

<https://1drv.ms/x/s!AqM9XZouV4HYkFnU5toyhvtB0zkz?e=nedLMS>

Detalhamento dos descritores moleculares:

<http://clri.res.in/subramanian/tools/drawing-tools/PaDEL-Descriptor/Descriptors.xls>

Notação SMILES de todas as moléculas utilizadas:

<https://1drv.ms/w/s!AqM9XZouV4HYkFwuwNjbua8wy2UC?e=yOehhN>