Gênisson dos Reis Santos

**Um estudo de QSAR sobre a interação de compostos
arilpiperazínicos com o receptor 5-HT_{2a} utilizando os métodos
PLS e ANN**

Dissertação apresentada ao Instituto de
Química de São Carlos da
Universidade de São Paulo como parte
dos requisitos para obtenção do título
de Mestre em Ciências.

Área de concentração: Físico-química

Orientador: Prof. Dr. Albérico Borges
Ferreira da Silva

Exemplar Revisado

O exemplar original encontra-se em
acervo reservado na Biblioteca do IQSC-USP

São Carlos, SP

2017

Às mulheres da minha vida: Genilma, Anna Raissa e Thiana.

Agradecimentos

Primeiramente agradeço a Deus Pai Celestial, criador de todas as coisas e de todos os mistérios da natureza ao nosso redor, os quais o seu fascínio exercido é tão grande sobre os homens que desprendemos grande parte de nosso tempo e de nossa existência na tentativa incessante de desvendá-los de uma forma lógica e racional, a qual chamamos de ciência.

Agradeço de todo coração a minha mãe, Genilma, que me criou, me educou, que sempre me ensinou a lutar, a persistir e a ser maior que as adversidades da vida. Agradeço a ela também que foi minha primeira professora e educadora que me orientou ao longo dessa minha jornada e que me orienta até hoje.

Agradeço também a minha irmã, Anna Raissa, que sempre esteve ao meu lado, sempre cuidou e zelou por mim, minha velha companheira de caminhada e que sempre acreditou muito em meu potencial, a qual eu sei que tem muito orgulho da pessoa que eu me tornei como adulto assim como eu tenho muito orgulho também dela.

Agradeço a minha namorada, Thiana, por todo apoio moral, por todos os dias que me desejou um dia abençoado, por todo carinho, amor e atenção e por sua participação direta na conclusão desse trabalho onde me auxiliou enormemente com a confecção das estruturas químicas de moléculas.

Ao senhor Célio Lipinski e a Dr^a. Aline Alves Oliveira, meus dois queridos colegas de laboratório que me aturaram dia após dia na labuta do lab, e me ensinaram tudo que eu sei de Linux, linhas de comando e foram também fundamentais para a confecção dessa dissertação de mestrado. Meus mais sinceros agradecimentos a vocês, muito obrigado.

Ao meu professor orientador, Dr. Albérico Borges Ferreira da Silva que acreditou no meu potencial como pesquisador e me acolheu em seu laboratório como aluno e como mais um de seus filhos adotivos nessa grande família que é o laboratório de química quântica do IQSC-USP.

Também agradeço aos meus demais colegas de laboratório que fizeram dessa minha estadia em São Carlos um dos momentos mais divertidos e gratificantes da minha vida na pessoas do senhor Regis, dona Aldinéia, senhorita Natiele, dona Júlia, senhor Guilherme, Laíse, Gabriel, Eriosvaldo e Anne. Muito obrigado a vocês.

Por fim meus agradecimentos ao Instituto de Química de São Carlos e a Universidade de São Paulo pela infraestrutura concedida e a agência de fomento CNPq pela bolsa de estudos de mestrado.

No Meio do Caminho
(Carlos Drummond de Andrade)

No meio do caminho tinha uma pedra
Tinha uma pedra no meio do caminho
Tinha uma pedra
No meio do caminho tinha uma pedra

Nunca me esquecerei desse acontecimento
Na vida de minhas retinas tão fatigadas
Nunca me esquecerei que no meio do caminho
Tinha uma pedra
Tinha uma pedra no meio do caminho
No meio do caminho tinha uma pedra.

Resumo

A depressão, também conhecida como distúrbio depressivo maior, é uma doença extremamente séria caracterizada por distúrbios afetivos que apresenta sintomas como tristeza, pessimismo, baixa autoestima, alterações nas funções vegetativas como sono e apetite e cognitivas como concentração, memória e atenção. A depressão afeta hoje cerca de 300 milhões de pessoas sendo essa a doença mais incapacitante do mundo, segundo dados da Organização Mundial da Saúde. O receptor humano 5-HT_{2a} tem sido associado a inúmeras condições neurológicas e moléculas ligantes seletivas a esse receptor podem apresentar potencial terapêutico no tratamento de distúrbios de comportamento, tais como a depressão. Com o propósito de contribuir para o planejamento de novos fármacos mais eficazes no tratamento da depressão realizou-se um estudo de QSAR (do inglês *Quantitative structure-activity Relationship*) com um conjunto de 106 compostos arilpiperazínicos utilizando os métodos de Mínimos Quadrados Parciais (PLS) e Redes Neurais Artificiais (ANN). O modelo de PLS foi obtido com 76 compostos no conjunto de treinamento e 11 compostos no conjunto teste ($r^2 = 0,749$ e $q^2 = 0,696$). Os testes de validação *leave-N-out*, randomização e detecção de *outliers* confirmaram a robustez e estabilidade dos modelos e demonstraram que os mesmos não foram gerados através de correlações ao acaso. O modelos de Redes Neurais Artificiais de Perceptron de Multicamadas (MLP-ANN) foi gerado utilizando funções de transferência *tansig-tansig*. O treinamento das redes demonstrou que o melhor modelo apresentava arquitetura 7-10-1. Tal modelo apresentou valores de erro quadrático médio (EQM) igual a 0,00964, desvio médio absoluto (DMA) igual a 0,0775 e $r^2_{\text{treinamento}}$, $r^2_{\text{validação}}$ e r^2_{teste} iguais a 0,794, 0,795 e 0,788, respectivamente. Os descritores dos modelos matemáticos gerados mostraram concordância com as propriedades moleculares dos compostos e os valores de atividade biológica preditos pelos modelos de PLS e ANN demonstraram que a interação dos descritores com a atividade biológica é satisfatoriamente descrita tanto por aproximações lineares quanto não-lineares.

Palavras-chave: Depressão, 5-HT_{2a}, arilpiperazinas, QSAR, PLS, ANN.

Abstract

Depression, major depressive disorder or clinical depression, is a serious mood disorder characterized by severe symptoms like emotional suffering and feeling miserable that affects vegetative functions as sleeping and eating and cognitive functions like concentration, memory and attention. The world mental health estimates that depression affects more than 300 million people worldwide and is the leading cause of world disability. The human receptor 5-HT_{2a} has been associated to many neurological dysfunctions. Molecules that selectively bind to this receptor present the potential to be used as an effective treatment of mental health disorders, such as depression. Our primary aim in this work is to contribute to the development of new drugs for the treatment of depression. Quantitative Structure-Activity Relationships (QSAR) studies with a series of 106 arylpiperazines that interact with the neuroreceptor 5-HT_{2a} were conducted employing the methods Partial Least Squares (PLS) and Artificial Neural Networks (ANN). The PLS model was obtained with a training set of 76 compounds and a 19 compounds test set ($r^2 = 0.749$, $q^2 = 0.696$). Leave-*N*-out analysis, biological activity randomization and outliers detection confirmed the robustness and stability of the models and showed that they were not obtained by chance correlations. Predictive models were also generated by Multilayer Perceptron Artificial Neural Network (MLP-ANN) trained with *tansig-tansig* transfer functions showed that the best results were obtained for a 7-10-1 architecture ($r^2_{\text{training}} = 0.794$, $r^2_{\text{validation}} = 0.795$ and $r^2_{\text{test}} = 0.788$). The descriptors used to construct the predictive models showed good agreement with the arylpiperazines molecular properties, and the biological activity predicted by the PLS and ANN methods suggested that the descriptor interactions can be described by a linear or a nonlinear approach as well.

Keywords: Depression, 5-HT_{2a}, arylpiperazines, QSAR, PLS, ANN.

LISTA DE FIGURAS

Figura 1 - Medicamentos utilizados no tratamento da esquizofrenia (1) e tuberculose (2).....	3
Figura 2 - Estrutura da molécula de serotonina.....	4
Figura 3 - Estrutura química do pindolol, o primeiro antagonista do receptor 5-HT _{1a} conhecido.....	6
Figura 4 - Fármacos inibidores da enzima monoamino oxidase (MAO).....	7
Figura 5 - Agentes antidepressivos tricíclicos (ATCs).....	7
Figura 6 - Antidepressivos heterocíclicos (AIRS).....	8
Figura 7 - Inibidores seletivos de recaptação da serotonina (ISRS).....	8
Figura 8 - Diagrama esquemático de uma rede neural artificial.....	19
Figura 9 - Funções de transferência utilizadas no treinamento de ANNs.....	20
Figura 10 - Distribuição do negativo do logaritmo da atividade biológica ao longo da variação da estrutura dos compostos estudados.....	26
Figura 11 - Coeficientes de regressão para o modelo PLS.....	34
Figura 12 - Detecção de <i>outliers</i> no modelo PLS.....	35
Figura 13 - Compostos identificados na análise de <i>outliers</i> com valores altos de alavancagem (composto 74) e valores fora dos limites dos resíduos de Student (compostos 22, 90 e 102). Os substituintes na região R ₁ estão evidenciados.....	35

Figura 14 - Testes de <i>leave-N-out</i> aplicados ao modelo PLS.....	37
Figura 15 - Teste de randomização aplicado ao modelo PLS.....	38
Figura 16 - Predição da atividade biológica pelo modelo PLS.....	39
Figura 17 – Representação da MLP-ANN selecionada.....	42
Figura 18 - Predição da atividade biológica pelo modelo ANN.....	42
Figura 19 - Compostos que apresentaram baixos valores de E_{LUMO2+}	45
Figura 20 – Compostos que apresentaram altos valores para o descritor HATS1v. Em vermelho estão destacados os substituintes volumoso nas posições R_1 e R_2 das moléculas.....	47
Figura 21 - Moléculas que apresentaram valores altos para o descritor $R3p+$. Destacado em vermelho estão átomos com alta polarizabilidade presentes nas posições $R1$, $R4$ e $R5$ destas moléculas.....	49
Figura 22 – Moléculas que apresentaram altos valores para o descritor $RDF110i$	50
Figura 23 – Moléculas que apresentaram altos valores para o descritor $SpMax6_Bhm$. A presença de unidades assimétricas de carga devido à átomos bastante eletronegativos nas extremidades das moléculas está evidenciada na cor vermelha.....	51
Figura 24 – Compostos que apresentaram valores apreciáveis para o descritor $Mor27p$. Em vermelho estão evidenciados os átomo com alta polarizabilidade presentes nas extremidades dos grupos substituintes dessas moléculas.....	53

Figura 25 – Compostos que apresentaram valores apreciáveis para o descritor JGI9. Em vermelho estão destacadas as regiões de transferência de carga entre átomos vizinhos.....55

LISTA DE TABELAS

Tabela 1 - Representação estrutural e valores de IC ₅₀ dos compostos derivados de arilpiperazina em estudo.....	22
Tabela 2 - Parâmetros estatísticos do modelo gerado pela técnica de algoritmo genético.....	29
Tabela 3 – Valores dos descritores selecionados para realizar a geração de modelos PLS e ANN (os compostos do conjunto teste encontram-se sublinhados).....	29
Tabela 4 - Variância percentual acumulada, SEV, SEC, PRESS, q ² e r ² dos modelos PLS obtidos para o conjunto treinamento.....	33
Tabela 5 - Resíduos de predição do conjunto teste pelo modelo PLS.....	39
Tabela 6 – Melhores modelos de MLP-ANN gerados para cada camada intermediária.....	41
Tabela 7 – Resíduos de predição do conjunto teste pelo modelo ANN.....	43
Tabela 8 – Descritores selecionados e suas definições.....	44

LISTA DE SIGLAS E ABREVIATURAS

3D-MoRSE	Representação Molecular de Estruturas baseadas em Difração Eletrônica
5-HT	5-Hidroxitriptamina
ANN	Redes Neurais Artificiais
ATC	Antidepressivos Tricíclicos
CRH	Hormônio Liberador de Corticotropina
DA	Dopamina
DFT	<i>Teoria do Funcional da Densidade</i>
DMA	Desvio Médio Absoluto
EQM	Erro Quadrático Médio
GA	Algoritmo Genético
GETAWAY	Geometria, Topologia e Pesos-Atômicos
GPCRS	Receptores acoplados à proteína G
HOMO	Orbital Molecular Ocupado de Maior Energia
IMAO	Inibidores de Monoaminoxidase
ISRS	Inibidores da Recaptação da Serotonina
LUMO	Orbital Molecular Desocupado de Menor Energia
MAO	Monoamino-Oxidase
MIM	Matriz de Influência Molecular
MLP-ANN	Redes Neurais Artificiais de Perceptron de Multicamadas
NE	Norepinefrina
PLS	Mínimos Quadrados Parciais
PRESS	Somas de Erros de Resíduos de Predição
QSAR	Relação Quantitativa Estrutura-Atividade
RDF	Função de Distribuição Radial
SEC	Erro Padrão de Calibração

SERT	Transportadores de Serotonina
SEV	Erro Padrão de Validação
TEC	Terapia Eletroconvulsiva

Sumário

1 INTRODUÇÃO	1
1.1 Depressão	1
1.2 A Hipótese Monoaminérgica	3
1.3 A serotonina e seus receptores	4
1.4 Os Antidepressivos	6
2. ESTUDOS DE QSAR	9
2.1 Requisitos do Conjunto de Dados	11
2.2 Cálculo de Descritores Moleculares	12
2.3 Seleção de Descritores	13
2.3.1 A matriz de correlação	13
2.3.2 Algoritmo genético	14
3. CONSTRUÇÃO DO MODELO DE QSAR	15
3.1 PLS	15
3.2 ANN	18
4 OBJETIVOS	21
4.1. Objetivo Geral	21
4.2. Objetivos Específicos	21
5. METODOLOGIA	22
6. RESULTADOS E DISCUSSÃO	28
6.1 Seleção de descritores	28
6.2 Geração dos Modelos QSAR	32
6.2.1 Análise de Mínimos Quadrados Parciais (PLS)	32
6.2.2 Análise de ANN	40
6.3 Análise dos Descritores Seleccionados	44

6.3.1 ELUMO2+	45
6.3.2 HATS1v	46
6.3.3 R3p+	48
6.3.4 RDF110i	49
6.3.5 SpMax6_Bhm	51
6.3.6 Mor27p	52
6.3.7 JGI9	54
7. CONCLUSÕES E PERSPECTIVAS	56
8. REFERÊNCIAS	57

1. INTRODUÇÃO

1.1. Depressão

Depressão, distúrbio depressivo maior ou depressão clínica é um distúrbio de humor comum, embora sério, que é fruto de uma interação complexa entre fatores psicológicos, sociais e biológicos. Pessoas que passaram por eventos adversos na vida (demissão, acidente grave, trauma psicológico) apresentam uma maior chance de desenvolver a depressão. A depressão pode, por sua vez, levar a mais stress e perda de ânimo, vindo a agravar a situação da pessoa deprimida. Os sintomas da depressão geralmente apresentam-se como alterações nas funções vegetativas (sono, apetite e peso), cognitivas (concentração, memória, atenção, tomada de decisões), comportamentais (interesse, motivação, prazer), somáticas (cefaleias, dores estomacais, tensão muscular) e de controle do impulso (suicídio, homicídio) de um indivíduo. No entanto, nem todos os indivíduos diagnosticados com depressão experimentam todos os sintomas, enquanto outros podem apresentar vários. Sintomas depressivos persistentes em adição a tristeza constante e baixa autoestima são requisitos para o diagnóstico da depressão.¹

Algumas formas de depressão são ligeiramente diferentes das outras, ou se desenvolvem sob circunstâncias particulares tais como: transtorno depressivo persistente, depressão perinatal, depressão psicótica, desordem afetiva temporária e transtorno bipolar. Sendo o último diferente da depressão, mas está incluído dada sua semelhança nos sintomas, ao passo que a pessoa com transtorno bipolar experimenta episódios de extrema baixa autoestima que obedecem aos critérios da depressão maior.²

Atualmente a depressão é a doença mais incapacitante do mundo, afetando mais de 300 milhões de pessoas ao redor do planeta. Estudos europeus apontaram que a depressão é responsável por 47% das aposentadorias precoces, 32% das licenças por problemas de saúde o que tem um auto impacto socioeconômico, uma vez que sintomas depressivos como

falta de atenção, perda de memória, e dificuldade de planejamento e tomada de decisões prejudicam a habilidade de trabalhar de um indivíduo.³ Nos casos mais severos a depressão pode levar até mesmo ao suicídio. Cerca de 800.000 pessoas morrem por suicídio a cada ano, sendo esta a segunda maior causa de mortes entre jovens de 15-29 anos de idade.⁴

Os gastos com a depressão e outros distúrbios de saúde mental estão em crescimento. De acordo com Koff Annan, ex-secretário geral das nações unidas, que abriu o seminário *The Global Crisis of Depression* em Londres em novembro de 2014, a depressão atinge hoje custos diretos e indiretos estimados em US\$ 800 bilhões ao redor do mundo.⁵ A Organização Mundial da Saúde estimou que a depressão seria responsável por uma grande parcela dos gastos com doenças, saindo de terceiro lugar em 2004 com 4,3% dos gastos totais para primeiro lugar em 2030 com cerca de 6,2% do total de gastos com saúde, seguida de isquemia cardíaca e acidentes de trânsito.

Embora existam tratamentos eficazes conhecidos para depressão (medicação, aliada a psicoterapia é efetiva em 60 a 80% dos casos), menos da metade das pessoas afetadas (em muitos países, menos de 10%) recebem tais tratamentos. As barreiras para um tratamento eficiente da depressão incluem falta de recursos, ausência de uma política de saúde pública adequada e o estigma social associado aos transtornos mentais. Neste contexto, ignorância e desinformação levaram à crença de que a depressão não seria uma doença, mas consistiria numa deficiência de caráter, a qual poderia ser superada com esforço próprio.²

Mesmo sob tratamento médico, aproximadamente 30% dos pacientes diagnosticados com depressão não respondem à terapia farmacológica, devido ao tempo de latência associado ao surgimento do efeito terapêutico e aos seus efeitos colaterais adversos como dores fortes de cabeça, náuseas, ansiedade, insônia, nervosismo, dentre outros, o que leva vários pacientes a abandonarem a terapia.⁶ Por esses motivos existe a necessidade de desenvolvimento de novos fármacos com melhores características, isto é, propriedades farmacológicas aprimoradas e efeitos colaterais reduzidos, para o tratamento do distúrbio depressivo maior.⁷

1.2. A Hipótese Monoaminérgica

Existem atualmente várias teorias que procuram explicar a fisiopatologia do distúrbio depressivo maior. Uma dessas teorias, conhecida como a hipótese monoaminérgica, propõe que a falta de uma ou mais monoaminas neurotransmissoras, serotonina (5-HT), noradrenalina (NE) e dopamina (DA), no sistema nervoso central dá origem ao desenvolvimento da depressão.⁸

Uma evidência clínica que serviu para sustentar esta hipótese foi a observação de que o uso crônico da roserpina (1), no tratamento da hipertensão e esquizofrenia, está caracterizadamente associado ao aumento do risco de aparecimento da depressão em cerca de 25% dos indivíduos. Esta evidência clínica, associada à elucidação do principal mecanismo de ação desta droga, que consiste na inibição do armazenamento de serotonina e norepinefrina nas vesículas das terminações nervosas, suportou a hipótese de que o efeito depressivo resulta da depleção de uma ou mais aminas biogênicas no cérebro. Em seguida, demonstrou-se que a isoniazida (2), utilizada no tratamento da tuberculose, aliviava a depressão em pacientes com tuberculose crônica e que esses efeitos resultavam de sua capacidade em prevenir a degradação das aminas biogênicas, inibindo a enzima monoamino-oxidase (MAO). Estas observações conduziram ao silogismo, que constituiu a base da hipótese monoaminérgica da depressão.⁹

Figura 1 - Medicamentos utilizados no tratamento da esquizofrenia (1) e tuberculose (2).

Contudo, outras hipóteses têm surgido recentemente. Sabe-se que o hipotálamo produz a substância CRH (*corticotropin-releasing hormone*), que

estimula a glândula pituitária e desencadeia a liberação de hormônios glucocorticóides, que são hormônios do estresse, como o cortisol. Por isso, essa hipótese é conhecida como hipótese do estresse e é sustentada por estudos com animais que mostram uma diminuição das substâncias que mantêm os neurônios saudáveis, causada por esses hormônios. Por outro lado, sabe-se que animais deprimidos apresentam um baixo crescimento de novos neurônios (neurogênese). Tais conhecimentos deram origem à teoria neurotrófica, que está de acordo com o que se conhece sobre os efeitos do estresse no cérebro, bem como com os resultados de estudos que indicam que o hipocampo e o córtex pré-frontal apresentam-se atrofiados em animais cronicamente deprimidos.¹⁰

1.3. A serotonina e seus receptores

5-Hidroxitriptamina, mais comumente conhecida como serotonina (Figura 2), é uma monoamina neurotransmissora encontrada por todo corpo humano. O 5-HT é sintetizado em uma pequena população de neurônios nos núcleos de raphe no mesencéfalo, que expressa a enzima Tryptophan hidroxilase. No entanto, a maior parte da serotonina encontrada no sangue é produzida por células enterocromafins, no trato gastrointestinal.

Figura 2 - Estrutura da molécula de serotonina.

Serotonina (3)

A serotonina se liga a mais de 14 receptores de proteínas, das quais a maioria são receptores acoplados a proteína G (*G-proteincoupled receptors*, GPCRs). A transdução de sinais através da membrana feita por um receptor desse tipo não é completamente entendida. Sabe-se que a proteína G inativa

existe ligada ao receptor em seu estado inativo. Uma vez que um ligante seja reconhecido, o receptor muda de conformação e ativa a proteína G, que irá dissociar a subunidade alfa ($G\alpha$) do dímero $G\beta\gamma$ e do receptor, dando continuidade à cascata de sinais celulares.¹¹ Acredita-se que essa única molécula está envolvida em inúmeros processos fisiológicos e cognitivos por todo o corpo incluindo: emoções, memória, sono, e regulação térmica.

Indícios mostram que os GPCRs existem num equilíbrio conformacional entre os estados biofísicos ativo e inativo, e que a interação com um ligante pode deslocar este equilíbrio na direção de um ou outro estado do receptor. Os ligantes conhecidos como agonistas deslocam o equilíbrio na direção do estado ativo; os agonistas inversos favorecem o estado inativo e os antagonistas são ligantes que não afetam este equilíbrio. Entretanto, ainda não se sabe como os estados ativo e inativo diferem entre si.¹²

Autoreceptores 5-HT estão amplamente distribuídos ao longo do cérebro, assim também como transportadores de serotonina (SERT), e são conhecidos por desempenhar um importante papel no controle dos níveis sinápticos. O bloqueio de receptores 5-HT_{1a}, 5-HT_{2a}, 5-HT_{1b}, 5-HT_{2c} com ou sem a inibição concomitante dos SERT rapidamente aumenta a concentração de serotonina no cérebro e conseqüentemente é capaz de desencadear uma rápida ação antidepressiva/anti-ansiedade.⁸

A despeito do surgimento de novas estratégias e hipóteses nesse campo de estudos, os receptores 5-HT têm sido alvo de intensas pesquisas nas últimas décadas, uma vez que muitos fármacos eficientes no tratamento da depressão e ansiedade agem sobre este neuromodulador, tanto na forma de inibidores da recaptação de 5-HT (ISRS), quanto agonistas ou antagonistas de seus subtipos de receptores, 5-HT_{1A}, 5-HT_{1B}, 5-HT_{2A} e 5-HT_{2C}.¹³ Assim, postula-se que a chave para o tratamento da depressão sejam os receptores 5-HT, que devem ser dessensibilizados para permitir a ação antidepressiva. A dessensibilização desses receptores é observada em ratos sob tratamento crônico com antidepressivos ou terapia eletroconvulsiva (TEC). Essa suposição é sustentada pelo aumento da concentração de 5-HT extracelular após a administração de antagonistas do receptor 5-HT_{1a}, tais como o pindolol (cuja estrutura é mostrada na Figura 3), juntamente com ISRSs, que têm diminuído o

tempo de latência e melhorado a eficácia do tratamento com esses antidepressivos.¹⁴

Figura 3 - Estrutura química do pindolol, o primeiro antagonista do receptor 5-HT_{1a} conhecido.

1.4. Os Antidepressivos

Os primeiros medicamentos antidepressivos foram descobertos por acaso quando observou-se que fármacos usados no tratamento da tuberculose apresentavam também a capacidade de melhorar o humor dos pacientes. Atualmente, os principais medicamentos antidepressivos utilizados na prática médica podem ser agrupados em três classes: inibidores de monoaminoxidase, antidepressivos tricíclicos e inibidores seletivos da recaptação de serotonina, explicados a seguir:

- Inibidores de Monoaminoxidase (IMAO): Os inibidores de MAO estão entre os primeiros fármacos utilizados no tratamento da depressão. A maioria destes agentes inibe não somente a MAO, mas também outras enzimas e interfere no metabolismo hepático de muitos fármacos. A MAO é uma enzima portadora de flavina, localizada primariamente na membrana externa das mitocôndrias e encontrada nos terminais nervosos, no fígado e em outros órgãos. Esta enzima oxidativa, inativa as amins biogênicas tais como norepinefrina, dopamina, serotonina, triptamina e tiramina, convertendo-as nos aldeídos correspondentes. Os primeiros inibidores da MAO a serem utilizados no tratamento da depressão foram a fenelzina (5) e a tranilcipromina (6).¹⁵

Figura 4 - Fármacos inibidores da enzima monoamino oxidase (MAO).

- Antidepressivos Tricíclicos (ATC): foram descobertos a partir da síntese de derivados dibenzazepínicos em 1951 e eram inicialmente empregados terapeuticamente por suas propriedades sedativas e analgésicas em animais. Durante a investigação clínica de isósteros fenotiazínicos observou-se que os novos derivados tinham efeito apreciável em pacientes deprimidos (Figura 4). Os ATCs, representados pela imipramina (7) e amitriptilina (8), são relativamente não-seletivos em suas ações, sendo predominantemente caracterizados como inibidores da recaptação de NE e 5-HT. Em geral, os metabólitos monometilados correspondentes (desipramina e nortriptilina) são mais potentes e mostram ação mais rápida que (7) e (8). Existem indícios de que a habilidade dos fármacos tricíclicos atuarem como inibidores do mecanismo de recaptação neuronal está associada às suas similaridades conformacionais com a NE.¹⁶

Figura 5 - Agentes antidepressivos tricíclicos (ATCs).

A partir de 1980, foram introduzidos novos antidepressivos de segunda geração, como a amoxapina (9) e maprotilina (10), estruturalmente relacionados aos tricíclicos clássicos, além da trazodona e nefazodona que, por apresentarem neurofarmacologia menos definida, são considerados atípicos. Estes fármacos possuem maior seletividade tendo efeito primário sobre a

recaptação de NE, 5-HT e DA. Entretanto, não diferem notavelmente dos agentes clássicos em termos de potência..¹⁷

Figura 6 - Antidepressivos heterocíclicos (AIRS).

- Inibidores Seletivos da Recaptação de Serotonina (ISRS): estão envolvidos no aumento da neurotransmissão de 5-HT_{1a}, como resultado da dessensibilização dos autoreceptores 5-HT somatodendríticos e terminais (exemplos: fluoxetina e paroxetina). Fármacos como (11), fluvoxamina (12), e citalopram (13) são exemplos de ISRS comerciais..¹⁸

Figura 7 - Inibidores seletivos de recaptação da serotonina (ISRS).

Os ISRS representam a classe mais prescrita pelos médicos, por apresentarem menor incidência de efeitos colaterais. Entretanto, aproximadamente 30% dos pacientes não respondem à medicação. Além disso, o tempo de latência até que esses medicamentos comecem a apresentar efeito terapêutico significativo é de 2 a 6 semanas, tempo considerado muito longo por haver possibilidade de suicídio..¹⁷

O processo de descoberta e desenvolvimento de fármacos é complexo, longo e de alto custo, tendo suas raízes profundamente ligadas às inovações

científicas e tecnológicas.¹⁹ A química medicinal é de reconhecido papel no processo de pesquisa e desenvolvimento de fármacos abrangendo diversas especialidades, tais como química orgânica, bioquímica, farmacologia, informática, biologia molecular e estrutural.⁹

No âmbito desse assunto nasce a motivação para este trabalho na busca de um maior entendimento das propriedades moleculares e eletrônicas relacionadas à afinidade de moléculas derivadas de arilpiperazina, fármacos em potencial, com receptores da família 5-HT, em especial o receptor 5-HT_{2a}. Para tal propõe-se um estudo de QSAR com emprego dos métodos PLS e ANN, na elucidação de um modelo matemático capaz de prever a atividade biológica desses compostos, cujos fundamentos serão apresentados a seguir.

2. ESTUDOS DE QSAR

Os estudos de Relação Quantitativa Estrutura-Atividade (QSAR) se tornaram ferramentas de grande utilidade para a indústria farmacêutica, aplicados na descoberta e otimização do desenvolvimento pré-clínico de fármacos, e na descrição e predição de propriedades de sistemas complexos de moléculas em diferentes ambientes.²⁰ O sucesso do QSAR deve-se bastante à elucidação da estrutura responsável por uma atividade biológica de interesse e a possibilidade de estimar as propriedades de um novo composto químico antes mesmo que ele tenha sido sintetizado.²¹

A premissa principal que fundamenta os estudos de QSAR é que as interações dos fármacos com seus alvos biológicos são determinadas por forças intermoleculares, ou seja, interações hidrofóbicas, polares, eletrostáticas e estéricas. Os fármacos que exercem seus efeitos biológicos interagindo com um alvo específico (uma enzima, um receptor, um ácido nucléico ou qualquer outra macromolécula biológica) devem ter uma estrutura tridimensional tal que os grupos funcionais e as propriedades de superfície sejam aproximadamente complementares ao sítio de interação. Desde que o sistema biológico seja mantido constante, a interação entre fármacos diferentes com o sítio ativo de um alvo biológico, bem como suas distribuições no sistema, dependem apenas

das estruturas químicas dos compostos. Se estas estruturas estão intimamente relacionadas, as diferenças em suas propriedades físico-químicas e assim as diferenças nas forças de interação, podem ser facilmente descritas de uma maneira quantitativa de modo que a variação correspondente nas atividades biológicas pode ser diretamente relacionada às variações nessas propriedades. Assim, todos os modelos quantitativos de relação estrutura-atividade são baseados na suposição de uma aditividade das contribuições de grupos aos valores da atividade biológica.²²

Existem muitos propósitos práticos de um estudo QSAR e estas técnicas são utilizadas amplamente em várias situações, como por exemplo:

- Predizer a atividade biológica e propriedades físico-químicas de maneira racional;
- Compreender e racionalizar os mecanismos de ação de uma série de compostos.

Por trás desses objetivos, as razões para se desenvolver estes modelos incluem:

- Economia no custo de desenvolvimento de produtos, por exemplo, em áreas farmacêuticas, pesticidas, produtos pessoais, etc;
- Redução (e em alguns casos substituição) de testes em animais, diminuindo o uso destes em pesquisas e obviamente dor e desconforto gerado a eles;
- Promoção de química verde, aumentando a eficiência de processos e evitando desperdícios.²³

O estudo de relações quantitativas estrutura-atividade geralmente envolve 3 passos: (1) coleta e, se possível, desenvolvimento de um conjunto teste de compostos; (2) a escolha adequada de descritores que podem relacionar apropriadamente a estrutura química à atividade biológica; e (3) aplicação de métodos estatísticos que correlacionem as mudanças estruturais com variações na atividade biológica. Por fim, a aplicação de testes de

validação dos modelos, são imprescindíveis para assegurar a confiabilidade e a utilidade dos mesmos. A obtenção de um modelo QSAR de boa qualidade com a habilidade de prever a atividade biológica de compostos químicos fora do conjunto de treinamento depende da qualidade das aproximações feitas e execução cuidadosa de cada uma dessas três etapas.²⁴

2.1. Requisitos do Conjunto de Dados

Uma importante condição em um estudo QSAR é a disponibilidade de uma série de compostos com um esqueleto comum entre si, que apresentem o mesmo mecanismo de ação.²⁵ Além disso, o conjunto de dados deve ser oriundo do mesmo protocolo de testes, e cuidado deve ser tomado com respeito à variabilidade Inter-laboratorial das medidas. Qualquer escolha incorreta dos dados pode comprometer a correlação apropriada entre estrutura e atividade.

Algumas regras de escolha de um bom conjunto de dados para se dar início a um estudo de QSAR são que a dosagem da resposta da relação deve ser regular, a afinidade (ou potência) reprodutível, a amplitude da variação da atividade deve conter duas ou mais ordens de magnitude do composto menos ativo para o mais ativo da série, a quantidade de compostos usados na construção do modelo QSAR deve ser suficientemente grande para assegurar estabilidade estatística, a atividade dos compostos deve estar suavemente distribuída ao longo da escala de atividades e os compostos selecionados para o conjunto teste devem apresentar uma diversidade estrutural suficiente para cobrir a escala química associada com a atividade biológica em estudo.²⁴

Quanto aos dados biológicos, os seguintes tipos de dados de atividade biológica podem ser usados em estudos QSAR, desde que estejam na escala correta: valores de atividade biológica *in vitro* (usando culturas de bactérias, fungos e outras, bem como órgãos isolados) e *in vivo* (atividades farmacodinâmicas e tóxicas de fármacos); dados de afinidade, como constantes de interação com um receptor ou substrato; constantes de velocidade, como associação/dissociação e constantes de Michaelis-Menten; constantes de inibição, especialmente valores de K_i e IC_{50} de diferentes

enzimas; parâmetros farmacocinéticos, como constantes de velocidade de absorção, parâmetros de distribuição, constantes de velocidade de degradação metabólica e constantes de velocidade de eliminação.²²

2.2. Cálculo de Descritores Moleculares

As estruturas químicas não contêm a informação relacionada à atividade biológica de maneira explícita. Esta informação deve ser extraída das estruturas na forma de descritores moleculares, que são parâmetros quantitativos característicos de cada molécula e fornecem informações sobre suas estruturas. Os descritores têm o papel de acentuar diferentes propriedades químicas implícitas na estrutura da molécula. Tais propriedades, que vão desde parâmetros físico-químicos, eletrônicos ou químico-quânticos a características geométricas e topológicas, podem ser então correlacionadas diretamente com a atividade. Tecnicamente, os descritores moleculares convertem a estrutura à forma de conjuntos de valores numéricos bem definidos para serem usados na análise estatística dos dados, representando várias propriedades moleculares que são consideradas importantes para explicar a atividade biológica.²⁷

Os descritores podem ser divididos em várias classes, podendo-se citar os eletrônicos, topológicos, geométricos e constitucionais, por exemplo. Eles possibilitam obter informações sobre a molécula como tamanho e forma, distribuição de cargas, interações intramoleculares, distribuições de energia, dentre muitas outras²⁴, tornando possível adquirir informações necessárias para o entendimento da relação entre a estrutura molecular com sua atividade.

Um dos métodos computacionais mais utilizados para que se dê início a etapa de cálculo de descritores é o método quântico DFT (do inglês *Density Function Theory*), que produz resultados satisfatórios a um custo computacional relativamente baixo. O método DFT se baseia na resolução da equação de *Schrödinger* para elétrons de sistemas moleculares nos quais o estado fundamental de energia de um átomo ou molécula é expresso em termos de sua densidade eletrônica. Através desse método, pode-se realizar o cálculo de propriedades de átomos e moléculas, como eletronegatividade,

polarizabilidade isotrópica, potencial químico, dentre outras. Diversas publicações mostraram que o DFT é um método eficaz para realizar o cálculo dos descritores químicos, que representam as estruturas moleculares de forma numérica, a serem utilizados em estudos de QSAR.^{28,29}

2.3. Seleção de Descritores

Muitos tipos de descritores de estruturas químicas estão disponíveis através de programas comerciais. A obtenção de um modelo estatisticamente robusto é muito dependente de como a qualidade da seleção dos descritores é capaz de codificar a variação da atividade de acordo com a estrutura. Quanto mais se conhece no nível molecular sobre o mecanismo biológico de ação dos compostos em estudo, mais se pode selecionar com eficiência dentro de um grupo vasto e abrangente de descritores moleculares. Algumas técnicas devem ser empregadas na seleção dos descritores para que possam ser definidos os melhores modelos QSAR. Dentre essas técnicas, podem ser utilizados para a seleção dos descritores a matriz de correlação e o algoritmo genético (GA).²⁴

2.3.1. A matriz de correlação

A matriz de correlação faz uso do coeficiente de correlação de Pearson (R_{ij}) para correlacionar as variáveis entre si, através da equação (1), na qual X_i e X_j correspondem aos valores das variáveis e n corresponde ao número de amostras.

$$R_{ij} = \frac{\sum X_i X_j - \sum X_i \sum X_j / n}{\sqrt{(\sum X_i^2 - (\sum X_i)^2 / n) \cdot (\sum X_j^2 - (\sum X_j)^2 / n)}} \quad (1)$$

No caso dos estudos de QSAR a matriz de correlação faz uma correspondência entre os valores dos descritores moleculares e a atividade biológica dos compostos em estudo. Valores muito baixos de R_{ij} indicam uma correlação muito baixa e valores muito altos indicam uma proximidade muito

grande, sendo considerados repetitivos por serem extremamente correlacionados, portanto baixos valores e altos valores de R_{ij} não são desejáveis. No entanto, a faixa de valores a serem considerados deve ser definida de acordo com o estudo a ser realizado. Em estudos de QSAR recomenda-se valores de correlação dos descritores com a atividade biológica entre 0,3 e 0,6.^{30,31}

2.3.2. Algoritmo genético

O algoritmo genético é uma técnica de otimização que mimetiza o fenômeno de seleção natural. A essência da seleção natural consiste no fato de, sob certas condições ambientais, espécies que melhor se adaptam são capazes de prevalecer nas gerações futuras, e que as melhores espécies são geradas através de cruzamentos ao acaso associados a mutações dos cromossomos das espécies sobreviventes.³² Por conta de sua simplicidade, flexibilidade, fácil manuseio, requisitos mínimos e perspectiva global, GA tem sido empregado com sucesso em vários estudos QSAR.

O aspecto diferenciado do algoritmo genético é a investigação simultânea de uma série de possíveis soluções, cada uma das quais explora diferentes regiões no espaço vetorial definido pelos descritores calculados. O primeiro passo de um processo GA típico é criar uma população de N indivíduos. Cada indivíduo codifica o mesmo número de descritores escolhidos ao acaso. A adaptação de cada indivíduo nessa geração é determinada. No segundo passo, uma fração de descendentes da próxima geração é produzida via cruzamento e o restante por mutação a partir dos ancestrais com base nos valores de adaptação deles. A nova prole contém então características dos dois ou de um de seus pais. O método também inclui elitismo, que por sua vez protege o indivíduo mais bem adaptado de cada geração de cruzamentos e de ser alterado durante a reprodução. O conteúdo genético desse indivíduo simplesmente passa adiante intacto para a próxima geração. Estes processos de seleção, cruzamento e mutação são repetidos até que todos os N pais na população sejam substituídos por seus descendentes. O valor de adaptação de cada membro desta nova geração é novamente avaliado, e o ciclo reprodutivo

continua até que 80 % das gerações apresentem o mesmo valor alvo de adaptação.²⁴

3. CONSTRUÇÃO DO MODELO DE QSAR

É também crítico que o método QSAR selecionado para desenvolver a correlação estrutura-atividade seja adequado. A relação entre um descritor molecular e a atividade biológica pode ser descrita de uma forma linear ou não-linear. É bastante difundido o uso de aproximações lineares como regressão linear multivariada (MLR) ou regressão de mínimos quadrados parciais (PLS) na construção do modelo de QSAR, já para a modelagem não-linear, o uso de redes neurais artificiais (ANN) oferece uma ótima alternativa.

3.1. PLS

O método de mínimos quadrados parciais (PLS) é uma técnica de análise de dados para o teste de relações teóricas dentro de um sistema de variáveis. O algoritmo fundamental do PLS baseia-se na correlação entre duas matrizes principais, X e Y, em que X representa os descritores e Y a atividade biológica. No método de PLS são obtidas as variáveis latentes (VLs), que são combinações lineares das variáveis originais das matrizes X e Y selecionadas de forma a descrever a máxima variância das amostras. As matrizes X e Y são decompostas simultaneamente em função da soma das VLs e das matrizes de erros que correspondem a parte não modelada, tal como nas equações 2 e 3:

$$X = TP' + E = \sum t_n p'_n + E \quad (2)$$

$$Y = UQ' + F = \sum u_n q'_n + F \quad (3)$$

Nessas equações T e U representam matrizes de *score* e P e Q matrizes de peso, E e F são as matrizes de resíduos, t_n e u_n são elementos das

matrizes de *score* e p_n e q_n elementos das matrizes de peso decompostos em n variáveis.

A correlação entre as matrizes X e Y é estabelecida através da correlação dos *scores* por meio de uma matriz B , que contém os coeficientes de regressão:

$$U = BT \quad (4)$$

desta maneira a matriz Y pode ser reescrita como:

$$Y = TBQ' + F \quad (5)$$

onde TB é o fator de regressão e F representa os erros da estimativa de Y .

A fim de se determinar os valores de erros de predição dos modelos PLS pode-se aplicar testes de validação cruzada com a exclusão de N compostos (*leave-N-out*). Neste teste de validação cruzada são retirados a cada passo n compostos do conjunto de calibração, que terão seus valores posteriormente preditos pelo modelo. Este procedimento é então repetido até que todas as amostras tenham sido retiradas para realizar a predição.

Do procedimento de validação cruzada obtém-se o parâmetro estatístico q^2 , que analisa a consistência do modelo de regressão e mensura a capacidade de predição de novos compostos.

$$q^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ival})^2}{\sum (y_{iexp} - y_{med})^2} \quad (6)$$

onde $\hat{y}_{ival}$ representa a predição da atividade sem a i -ésima amostra;

y_{iexp} : corresponde ao valor experimental da atividade da i -ésima amostra;

e y_{med} é valor médio experimental da atividade.

Um parâmetro importante utilizado para se atestar um bom modelo de PLS é o valor do coeficiente de correlação (r) o qual determina o valor mensurado da atividade biológica e o valor predito pelo modelo matemático gerado. Outro parâmetro utilizado na avaliação de um modelo de PLS é o valor

de r^2 , que diz respeito a fração da variabilidade total do sistema estudado que é explicada pelo modelo. Assim, quanto mais próximo o valor de r^2 for de 1, maior a capacidade do modelo de explicar as variações da variável em estudo em torno da sua média.

$$r^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - y_{med})^2} \quad (7)$$

onde $\hat{y}_{ical}$ corresponde à previsão da atividade da i -ésima amostra utilizando o modelo com todas as amostras.

A análise dos parâmetros estatísticos gerados pelos modelos de PLS permite definir um número ótimo de variáveis latentes. Recomenda-se que este procedimento seja feito buscando-se os menores valores dos erros de predição.³³ Dentre os parâmetros ligados aos erros de predição, os três principais são: a soma dos quadrados dos erros de predição (do inglês PRESS - *Prediction Residues Error Square Sum*), erro padrão de validação (do inglês SEV - *Standard Error of Validation*) e erro padrão de calibração (do inglês SEC - *Standard Error of Calibration*), calculados pelas equações (8), (9) e (10), respectivamente.

$$PRESS = \sum (y_{iexp} - \hat{y}_{ival})^2 \quad (8)$$

$$SEV = \left[\frac{PRESS}{I} \right]^{\frac{1}{2}} \quad (9)$$

$$SEC = \left[\frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{I - a - 1} \right]^{\frac{1}{2}} \quad (10)$$

O teste F fornece as razões entre as variabilidades explicadas e não explicadas do modelo. É indicado que os valores obtidos de teste F sejam altos indicando que o modelo é capaz de explicar suas próprias variações. O teste F é calculado à partir da seguinte equação:

$$Teste F = 1 - \frac{\sum(y_{iexp} - y_{med})^2 - (y_{iexp} - \hat{y}_{ical})^2}{\sum(y_{iexp} - \hat{y}_{ical})^2} \frac{(I - 1)}{VL} - 1 \quad (11)$$

em que I é o número de amostras e VL o número de variáveis latentes.

De forma resumida, pode-se afirmar que um bom modelo de PLS deve apresentar valores de r^2 e q^2 o mais próximo possível de 1, os maiores valores possíveis de Teste-F e os menores valores possíveis de *PRESS*, *SEV* e *SEC*.³⁴

Testes adicionais de validação também podem ser aplicados ao modelo PLS para verificar sua robustez, como a detecção de *outliers*, que identifica a presença de amostras atípicas e randomização, que identifica correlações ao acaso entre a variável dependente e os descritores.

3.2. ANN

Redes neurais artificiais são um sistema de processamento inspirado no sistema nervoso biológico. Elas consistem de um grupo de unidades chamadas de “neurônios” análogos aos neurônios nervosos. Cada neurônio está conectado aos outros por meio de pesos, de modo que a relação entre as variáveis de entrada e de saída podem ser determinadas. As ANNs quando empregadas em computadores não são programadas para executar uma tarefa específica, ao invés disso, elas são treinadas com base em conjuntos de dados até que aprendam os padrões apresentados por eles. Uma vez treinadas, novos padrões são apresentados a elas para previsão e classificação.³⁵

O método de redes neurais artificiais tem sido amplamente utilizado em uma vasta gama de aplicações. Estas incluem reconhecimento de padrões, simulação, previsão, automação, dentre inúmeras outras áreas de aplicação. Atualmente, pesquisadores tem dado preferência às ANN como modelo de previsão em detrimento de métodos estatísticos clássicos. Este interesse crescente pode ser explicado por três propriedades básicas das ANN: o primeiro é a ausência da necessidade de qualquer suposição prévia, a segunda é a extrapolação dos históricos gravados na geração de prognósticos e a última consiste na solução sucessiva de problemas complexos não lineares. As ANN

são amplamente aceitas como uma tecnologia que oferece uma estratégia alternativa no ataque de problemas complexos em situações atuais. A vantagem da ANN em relação aos outros métodos é a sua habilidade de modelar um problema de multivariáveis através da capacidade de extrair as relações implícitas não-lineares existentes dentro desse conjunto de variáveis via “aprendizado” a partir de um conjunto de treinamento.⁶

A rede mais comumente utilizada em estudos de QSAR é a MLP-ANN (do inglês *Multilayer Perceptron Artificial Neural Network*), que é do tipo perceptron de multicamadas. Ela é constituída pelas camadas de entrada, camadas intermediárias e camada de saída, nas quais as informações passam pelas camadas de entrada até atingirem a camada de saída. Cada camada possui um certo número de neurônios que são os elementos básicos de processamento da ANN. Cada neurônio é uma unidade de operação na ANN e está conectado à outros neurônios por meio de valores de pesos. O processo começa com a soma dos pesos de ativação de outros neurônios através das suas conexões de entrada, então, os pesos das somas são passados através de uma função de ativação, comumente chamadas de funções de transferência, e este valor ativado é o *output* do neurônio. Para estudos de QSAR, a camada de entrada é tida como os descritores e a camada de saída é o valor predito da atividade biológica.

Figura 8 - Diagrama esquemático de uma rede neural artificial.

As funções de transferência são capazes de conferir a não-linearidade a uma ANN, atuando sobre as entradas que chegam aos neurônios. Existem

diferentes funções de transferência, tais como as funções *logsig* (função de transferência logarítmica sigmoidal) e *tansig* (função de transferência tangente hiperbólica sigmoidal). Além dessas funções, existe também uma função linear representada pela *purelin* (função de transferência linear). A Figura 9 mostra uma representação esquemática das funções de transferência mencionadas.

Figura 9 – Funções de transferência utilizadas no treinamento de ANNs.

Para avaliar o desempenho de uma ANN, pode-se analisar os valores dos coeficientes estatísticos obtidos, tais como o coeficiente de determinação (r^2), erro quadrático médio (EQM) e desvio médio absoluto (DMA).^{36,37}

$$\text{EQM} = \frac{1}{N} \sum_{i=1}^N (y_{ical} - y_{iexp})^2 \quad (12)$$

$$\text{DMA} = \frac{1}{N} \sum_{i=1}^N |y_{ical} - y_{iexp}| \quad (13)$$

À luz dos métodos estatísticos apresentados faz-se possível a realização de estudos de QSAR para maior compreensão e entendimento da interação de compostos arilpiperazínicos com o receptor 5-HT_{2a}, no propósito de auxiliar no desenvolvimento de novos fármacos para o tratamento da depressão.

4. OBJETIVOS

4.1. Objetivo Geral

Desenvolver modelos QSAR capazes de prever a atividade biológica de uma série de compostos derivados de arilpiperazina frente ao receptor 5-HT_{2a}, a partir de descritores eletrônicos e moleculares.

4.2. Objetivos Específicos

- Desenhar e otimizar a estrutura molecular de uma série de 106 derivados de arilpiperazina.
- Obter e selecionar os descritores eletrônicos e moleculares que apresentem maior influência sobre atividade biológica dos compostos em estudo.
- Gerar um modelo matemático quantitativo capaz de prever a atividade biológica dos compostos utilizando os métodos de Redes Neurais Artificiais (ANN) e Mínimos Quadrados Parciais (PLS).
- Validar e selecionar os modelos preditivos estatisticamente mais bem-sucedidos.
- Analisar o significado químico do modelo gerado.

5. METODOLOGIA

Para que se desse início aos estudos de QSAR selecionou-se um conjunto de 106 compostos arilpiperazínicos sintetizados por Seo e colaboradores (SEO et al., 2011)⁷, ilustrados na Tabela 1, com atividades biológicas comprovadas por medidas de IC₅₀ frente ao receptor 5HT_{2a}. Arilpiperazina é o fragmento chave de um grande número de compostos ativos responsáveis por vários efeitos farmacológicos sendo os derivados de cadeia longo a classe de arilpiperazinas mais estudadas. Arilpiperazinas são encontradas como ligantes nos receptores de serotonina, especialmente os receptores 5-HT_{1a} e 5-HT_{2a}. Sua estrutura geral contém uma cadeia alquílica ligada um átomo de N da porção piperazínica e uma amina terminal em um fragmento imida.

Tabela 1 - Representação estrutural e valores de IC₅₀ dos compostos derivados de arilpiperazina em estudo

Estrutura I

Composto	R1	R2	R3	n	IC ₅₀ (nM)	pIC ₅₀
1	Ph	Me	Me	1	31,0	7,508638
2	Ph	Et	Me	1	27,0	7,568636
3	Ph	Pr	Me	1	39,0	7,408935
4	Ph	Me	Me	2	18,6	7,730487
5	Ph	Et	Me	2	48,0	7,318759
6	Ph	Pr	Me	2	21,0	7,677781
7	Ph	Ph	Me	2	14,2	7,847712
8	Pr	Me	Me	2	26,0	7,585027
9	c-Pentyl	Me	Me	2	55,0	7,259637
10	2-OMe-Ph	Me	Me	2	10,3	7,987163
11	2-OMe-Ph	Pr	Me	2	139	6,856985
12	4-OMe-Ph	Me	Me	2	20,0	7,698970
13	4-OMe-Ph	Pr	Me	2	5,17	8,286509

14	4-OMe-Ph	Ph	Me	2	20,0	7,886057
15	2-F-Ph	Me	Me	2	16,8	7,774691
16	2-F-Ph	Pr	Me	2	21,0	7,677781
17	2-F-Ph	Ph	Me	2	19,5	7,709965
18	2-Cl-Ph	Me	Me	2	30,0	7,522879
19	4-Cl-Ph	Pr	H	2	70,9	7,149354
20	2-CF ₃ -Ph	Me	Me	2	58,0	7,236572
21	y	Me	Me	2	8,63	8,063989
22	y	Pr	Me	2	106	6,974694

Estrutura II

Composto	R1	R2	R3	R4	R5	IC ₅₀ (nM)	pIC ₅₀
23	Ph	Me	Me	Me	Me	91,0	7,04095
24	Ph	Pr	Me	Me	Cl	154	6,81247
25	Ph	Pr	Me	Me	Me	75,0	7,12493
26	Ph	i-Pr	Me	Me	Me	105	6,97881
27	Ph	Pr	H	Me	Me	133	6,87614
28	i-Bu	Me	Me	Me	Cl	45,5	7,34198
29	i-Bu	Me	Me	Me	Me	217	6,66394
30	c-Pentyl	Me	Me	Me	Cl	80,6	7,09366
31	Ph	Ph	Me	Me	Me	29,4	7,53165
32	Ph	Ph	Me	Me	Cl	13,9	7,85698
33	Ph	Ph	H	Me	Me	13,2	7,87942
34	Ph	Ph		Me	Cl	12,3	7,91009
35	2-OMe-Ph	Me	Me	Me	Cl	12,8	7,89279
36	2-OMe-Ph	Me	Me	Me	Me	144	6,84163
37	2-OMe-Ph	Pr	Me	Me	Cl	171	6,76700
38	2-OMe-Ph	Pr	Me	Me	Me	253	6,59687
39	4-OMe-Ph	Me	Me	Me	Cl	23,1	7,63638
40	4-OMe-Ph	Me	Me	Me	Me	62,7	7,20273
41	4-OMe-Ph	Pr	Me	Me	Cl	15,2	7,81815
42	4-OMe-Ph	Pr	Me	Me	Me	70,8	7,14996
43	4-OMe-Ph	Ph	Me	Me	Cl	21,0	7,67778
44	4-OMe-Ph	Ph	Me	Me	Me	24,0	7,61978
45	4-OMe-Ph	Ph	H	Me	Me	19,0	7,72124

46	2-F-Ph	Me	Me	Me	Cl	40,0	7,39794
47	2-F-Ph	Me	Me	Me	Me	64,8	7,18842
48	4-F-Ph	Me	Me	Me	Cl	38,0	7,42021
49	4-F-Ph	Me	Me	Me	Me	118	6,92848
50	4-F-Ph	Pr	Me	Me	Cl	54,0	7,26760
51	4-F-Ph	Pr	Me	Me	Me	31,0	7,50863
52	4-F-Ph	Pr	H	Me	Me	141	6,85078
53	4-F-Ph	Ph	Me	Me	Cl	12,2	7,91364
54	4-F-Ph	Ph	Me	Me	Me	27,8	7,55909
55	2-Cl-Ph	Me	Me	Me	Cl	29,0	7,53760
56	2-Cl-Ph	Me	Me	Me	Me	80,0	7,09691
57	2-Cl-Ph	Pr	Me	Me	Me	90,0	7,04575
58	3-Cl-Ph	Me	Me	Me	Me	96,8	7,01412
59	4-Cl-Ph	Me	Me	Me	Me	67,0	7,17392
60	4-Cl-Ph	Pr	H	Me	Cl	20,0	7,69897
61	4-Cl-Ph	Pr	H	Me	Me	32,4	7,48945
62	3,5-diOMe-Ph	Me	Me	Me	Me	32,0	7,49485
63	y	Me	Me	Me	Cl	28,0	7,55284
64	y	Me	Me	Me	Me	40,6	7,39147
65	y	Pr	Me	Me	Cl	47,0	7,32790
66	y	Pr	Me	Me	Me	241	6,61798
67	y	Ph	Me	Me	Me	19,0	7,72124

Estrutura III

Composto	R1	R2	R3	R4	R5	IC ₅₀ (nM)	pIC ₅₀
68	Ph	Me	Me	Me	Me	248	6,606249
69	Ph	Pr	Me	Me	Cl	201	6,696804
70	Ph	Ph	Me	Cl	Cl	24,0	7,619789
71	Ph	Ph	Me	Me	Cl	57,0	7,244125
72	Ph	Ph	Me	Me	Me	58,6	7,232102
73	n-Pr	Me	Me	Me	Me	840	6,075772
74	i-Bu	Me	Me	Me	Me	1011	5,995249
75	c-Pentyl	Me	Me	Me	Cl	97,6	7,01055
76	2-OMe-Ph	Me	Me	Cl	Cl	52,1	7,283162
77	2-OMe-Ph	Me	Me	Me	Cl	370	6,431798
78	4-OMe-Ph	Me	Me	Me	Cl	56,0	7,251812

79	4-OMe-Ph	Me	Me	Me	Me	175	6,75721
80	4-OMe-Ph	Pr	Me	Cl	Cl	63,1	7,199971
81	4-OMe-Ph	Pr	Me	Me	Cl	582	6,235077
82	4-OMe-Ph	Pr	Me	Me	Me	126	6,899629
83	4-OMe-Ph	Ph	Me	Me	Cl	170	6,769551
84	4-OMe-Ph	Ph	Me	Me	Me	185	6,732828
85	3,4-diOMe-Ph	Me	Me	Me	Me	204	6,689307
86	2-F-Ph	Me	Me	Cl	Cl	104	6,982967
87	2-F-Ph	Me	Me	Cl	Cl	197	6,705534
88	2-F-Ph	Me	Me	Me	Me	579	6,237321
89	4-F-Ph	Me	Me	Cl	Cl	254	6,595166
90	4-F-Ph	Me	Me	Me	Me	78,3	7,106238
91	4-F-Ph	Pr	Me	Cl	Cl	196	6,707744
92	4-F-Ph	Pr	Me	Me	Me	327	6,485452
93	4-F-Ph	Ph	Me	Me	Me	71,0	7,148742
94	4-F-Ph	Ph	Me	Cl	Cl	94,0	7,026872
95	4-F-Ph	Ph	Me	Me	Cl	26,7	7,573489
96	2-Cl-Ph	Me	Me	Me	Me	318	6,497573
97	3-Cl-Ph	Me	Me	Me	Cl	182	6,739690
98	3-Cl-Ph	Me	Me	Me	Me	275	6,560509
99	4-Cl-Ph	Pr	H	Cl	Cl	177	6,753009
100	4-Cl-Ph	Pr	H	Me	Cl	107	6,970616
101	4-Cl-Ph	Pr	Me	Me	Me	147	6,832683
102	y	Me	Me	Cl	Cl	9,86	8,006123
103	y	Me	Me	Me	Cl	127	6,896196
104	y	Me	Me	Me	Me	232	6,635449
105	y	Pr	Me	Me	Cl	237	6,625252
106	y	Pr	Me	Me	Me	294	6,531653

Para que se tenha um bom estudo de QSAR, a faixa de variação de atividade biológica deve ser de pelo menos duas ordens de magnitude em unidades de concentração molar e estes valores são transformados na forma logarítmica para esse cálculo^{38,39}. Portanto, para que fosse feita uma seleção adequada de um grupo de compostos da literatura, fez-se necessário verificar a faixa de variação de atividade biológica, que para esses compostos se deu em termos de IC₅₀ (constante de dissociação do inibidor). Como pode ser visto na Figura 10, as atividades desses compostos apresentam uma boa distribuição dos valores de pIC₅₀ ao longo da variação das estruturas moleculares, entre 6,0 e 8,3, uma variação de cerca de 2,3 unidades logarítmicas nessa faixa de atividades biológicas. O que sugere que as modificações estruturais dos

compostos em questão estão intimamente relacionadas à variação da atividade biológica frente ao receptor 5HT_{2A}, dando indício de esse ser um conjunto de moléculas satisfatório para realização de um estudo de QSAR.

Figura 10 - Distribuição do negativo do logaritmo da atividade biológica ao longo da variação da estrutura dos compostos estudados.

As estruturas tridimensionais dos compostos foram construídas e otimizadas utilizando pacote Gaussian09⁴⁰. Os cálculos de otimização de geometria e frequências vibracionais foram conduzidos à luz da Teoria do Funcional de Densidade utilizando o funcional B3LYP⁴¹ e a função de base 6-31g++(d,p)⁴².

Dos resultados obtidos a partir dos cálculos de otimização foram extraídos os descritores eletrônicos polarizabilidade, momento de dipolo, energia da molécula e os orbitais de fronteira HOMO (do inglês *Highest Occupied Molecular Orbital*) e LUMO (do inglês *Lowest Unoccupied Molecular Orbital*). A partir das estruturas otimizadas desses compostos foram gerados descritores moleculares com o auxílio dos programas Dragon⁴³ e Padel⁴⁴. Com o objetivo de realizar uma seleção prévia e reduzir o grande número de descritores obtidos, realizou-se a análise da matriz de correlação entre cada um dos descritores e os valores de pIC₅₀ dos compostos arilpiperazínicos em estudo, onde aqueles descritores que apresentaram um valor de coeficiente de correlação (*r*) menores que 0,3 (em módulo) foram diretamente excluídos.

Dando seguimento aos estudos de QSAR iniciou-se a implementação da técnica de algoritmo genético para seleção de variáveis utilizando o programa BuildQSAR.⁴⁵ A seleção via algoritmo genético foi realizada variando-se o número de gerações, descritores por modelo e o índice de correlação entre variáveis. A escolha dos melhores modelos gerados via GA foi baseada nos valores de q^2 e r^2 .

Uma vez selecionados os descritores foram realizadas análises de PLS, com o auxílio do software Pirouette 3.10⁴⁶. Foram estudados tanto o conjunto de todos os 106 compostos como o conjunto de treinamento (76 compostos). Os parâmetros estatísticos q^2 , r^2 , PRESS, SEC e SEV foram utilizados como critério na seleção do melhor modelo preditivo e foram conduzidos procedimentos de validação como detecção de *outliers*, testes de *leave-N-out* e randomização da atividade biológica para verificação da robustez, estabilidade e ausência de correlações geradas ao acaso. Todos os testes de validação foram realizados no software QSAR Modeling⁴⁷.

Um modelo preditivo também foi gerado a partir do método de Redes Neurais Artificiais, em que o programa Matlab⁴⁸ foi utilizado para o treinamento das redes. Nas análises de ANN o algoritmo utilizado foi o de retropropagação com os valores dos descritores escalonados de 0,1 à 0,9 de acordo com a Equação 14, utilizando as funções de transferência do tipo tansig-tansig. A escolha do melhor modelo de ANN foi baseada na análise dos erros EQM e DMA e nos parâmetros estatísticos r^2 e q^2 .

$$i' = 0,8 \frac{i - i_{min}}{i_{max} - i_{min}} + 0,1 \quad (14)$$

6. RESULTADOS E DISCUSSÃO

6.1. Seleção de descritores

Após concluídos os cálculos de otimização de geometria e frequências vibracionais foram obtidos um total de 13 descritores eletrônicos, incluindo propriedades como momento de dipolo, energia total da molécula, polarizabilidade e energia dos orbitais de fronteira HOMO e LUMO, dentre outros. Com o auxílio dos programas Dragon e Padel gerou-se um total de 3947 descritores moleculares, dos quais 1191 oriundos do Dragon e 2756 oriundos do Padel.

Uma pré-seleção dos descritores eletrônicos e moleculares obtidos foi realizada baseando-se no valor do coeficiente da matriz de correlação (r) para cada descritor em relação aos valores de pIC_{50} . Os descritores que apresentaram baixa correlação com a atividade biológica (valores de r menores que 0,3 em módulo) foram excluídos, restando 571 descritores.

Dando seguimento ao trabalho foi iniciada uma etapa mais minuciosa de seleção via algoritmo genético utilizando o software BuildQSAR, com o intuito de evidenciar os descritores com maior relevância para a construção de modelos matemáticos. Variando o número de gerações de 500 até 10000, número de descritores de 4 a 10 e o critério de corte para a correlação entre eles, foram gerados um total de 700 modelos. Tomando como critério de avaliação os valores de r^2 e q^2 foram selecionados 93 modelos com os melhores valores frente essas variáveis.

Dos modelos gerados observou-se que o número ótimo de descritores era igual a 7, uma vez que essa era a quantidade mínima de descritores que apresentavam bons parâmetros estatísticos, enquanto que modelos com mais descritores, além de não melhorarem tais parâmetros, ainda se mostraram de difícil interpretação química.

Os melhores modelos foram então submetidos a análises prévias de PLS. O modelo que se destacou frente aos parâmetros estatísticos coeficiente de correlação (r), desvio padrão (s), teste - F (F) e o coeficiente de correlação

cruzada (q^2) continha os seguintes descritores: HATS1v, R3p+, RDF110i, SpMax6_Bhm, JGI9, E_{LUMO} 2+ e Mor27p. Os parâmetros estatísticos do melhor modelo gerado pela técnica de seleções de variáveis via algoritmo genético se encontram na Tabela 2 e os valores dos descritores selecionados na geração de modelos através das técnicas de PLS e de ANN estão apresentados na Tabela 3.

Tabela 2 - Parâmetros estatísticos do modelo gerado pela técnica de algoritmo genético

Descritores	r^2	s	F	q^2	PRESS	SEV
HATS1v, R3p+, RDF110i, SpMax6_Bhm, JGI9, E _{lumo +2} e Mor27p	0,741	0,255	40,02	0,700	0,274	0,265

Tabela 3 – Valores dos descritores selecionados para realizar a geração de modelos PLS e ANN (os compostos do conjunto teste encontram-se sublinhados)

Composto	E _{LUMO2+} (eV)	JGI9	SpMax6_Bhm	Mor27p	HATS1v	R3p+	RDF110i
1	-0,77335	0,008534	3,375673	-0,212	0,071	0,100	13,39038
2	-0,77634	0,008218	3,381654	-0,179	0,071	0,089	21,18630
3	-0,77416	0,007924	3,385817	-0,170	0,073	0,094	15,89026
4	-0,72355	0,007107	3,386365	-0,212	0,075	0,119	17,75396
5	-0,77335	0,006853	3,390046	-0,242	0,078	0,115	22,17331
6	-0,75457	0,006617	3,392726	-0,189	0,072	0,107	30,30286
7	-0,95648	0,005379	3,420272	-0,363	0,080	0,105	28,76481
8	-0,56328	0,008594	3,252288	-0,085	0,072	0,125	21,66612
9	-0,56355	0,007344	3,374710	0,122	0,069	0,121	17,26475
10	-0,69090	0,007306	3,387479	-0,227	0,072	0,108	25,05204
11	-0,68709	0,006835	3,393103	-0,132	0,069	0,097	34,16887
12	-0,71022	0,007729	3,390829	-0,237	0,067	0,115	19,33208
13	-0,71648	0,007408	3,394301	-0,281	0,063	0,101	22,48548
14	-0,37606	0,006188	3,433281	-0,444	0,071	0,094	29,80789
15	-0,71784	0,007567	3,387337	-0,199	0,075	0,120	21,68928
16	-0,73280	0,007063	3,393056	-0,196	0,070	0,104	26,90777
17	-0,99866	0,005809	3,422646	-0,317	0,079	0,100	24,37261
18	-0,72246	0,007567	3,391980	-0,148	0,077	0,116	22,51610
19	-0,76546	0,005729	3,397924	-0,162	0,079	0,110	22,77538
20	-0,71185	0,008125	3,391689	-0,244	0,072	0,112	29,11171

21	-0,70233	0,007159	3,395004	-0,184	0,067	0,112	18,35719
22	-0,70995	0,006892	3,395155	-0,222	0,064	0,094	21,03933
23	-0,60164	0,007107	3,152994	-0,300	0,063	0,061	22,75719
24	-0,62940	0,006617	3,306382	-0,206	0,063	0,072	34,83709
25	-0,59620	0,006617	3,221346	-0,187	0,056	0,057	36,29432
26	-0,59484	0,006617	3,221596	-0,156	0,054	0,056	35,28783
27	-0,53525	0,005625	3,209840	-0,192	0,059	0,056	32,53653
28	-0,37715	0,009050	3,263994	-0,128	0,060	0,082	20,97617
29	-0,18259	0,009050	3,137376	-0,101	0,053	0,064	25,21920
30	-0,36926	0,007344	3,287041	0,043	0,061	0,090	21,55489
31	-0,94151	0,005379	3,389137	-0,417	0,069	0,058	28,26793
32	-0,96192	0,004436	3,372918	-0,406	0,072	0,058	25,57932
33	-0,97036	0,004436	3,372918	-0,406	0,072	0,058	25,57932
34	-0,88682	0,004117	3,62657	-0,461	0,077	0,073	28,44438
35	-0,64681	0,007306	3,294048	-0,345	0,063	0,079	25,86521
36	-0,62505	0,007306	3,153191	-0,33	0,057	0,057	25,87178
37	-0,63022	0,006835	3,306518	-0,25	0,061	0,067	36,59682
38	-0,60872	0,006835	3,221398	-0,277	0,056	0,055	38,03780
39	-0,61607	0,007729	3,298561	-0,394	0,063	0,082	22,03579
40	-0,59675	0,007729	3,165431	-0,333	0,057	0,061	21,43638
41	-0,60899	0,007408	3,307606	-0,373	0,061	0,069	30,67857
42	-0,58940	0,007408	3,222654	-0,384	0,056	0,058	30,47672
43	-0,34286	0,006188	3,426616	-0,622	0,070	0,070	25,48782
44	-0,75403	0,006188	3,412658	-0,553	0,064	0,058	28,21677
45	-0,75022	0,005461	3,398518	-0,314	0,066	0,058	27,25119
46	-0,65226	0,007567	3,293917	-0,349	0,070	0,085	20,15899
47	-0,62858	0,007567	3,153101	-0,305	0,063	0,060	19,48855
48	-0,66015	0,007306	3,297511	-0,333	0,071	0,085	21,43407
49	-0,63049	0,007306	3,160556	-0,313	0,063	0,061	23,37630
50	-0,65389	0,006621	3,307343	-0,258	0,067	0,073	28,90511
51	-0,62015	0,006621	3,222190	-0,331	0,061	0,058	28,16371
52	-0,02041	0,005729	3,210533	-0,318	0,068	0,057	27,75322
53	-1,03730	0,006149	3,424495	-0,498	0,075	0,072	29,60248
54	-1,02179	0,006149	3,409978	-0,442	0,069	0,058	29,39573
55	-0,67838	0,007567	3,301588	-0,302	0,072	0,085	19,84059
56	-0,65525	0,007567	3,255883	-0,281	0,064	0,061	19,66074
57	-0,64654	0,007063	3,260652	-0,283	0,062	0,058	23,46952
58	-0,64545	0,007567	3,240613	-0,235	0,069	0,077	20,78483
59	-0,68682	0,007306	3,235926	-0,267	0,066	0,062	21,81366
60	-0,72927	0,005729	3,316201	-0,324	0,078	0,073	21,87657
61	-0,69716	0,005729	3,268618	-0,281	0,070	0,067	23,12380
62	-0,52055	0,007027	3,162576	-0,364	0,054	0,057	27,38026
63	-0,51375	0,007159	3,308674	-0,212	0,060	0,083	20,62379
64	-0,48491	0,007159	3,221156	-0,182	0,053	0,057	22,94195
65	-0,50586	0,006892	3,311202	-0,349	0,058	0,067	28,96815

<u>66</u>	-0,47593	0,006892	3,245395	-0,256	0,052	0,056	33,19517
67	-0,73852	0,005408	3,451794	-0,358	0,059	0,055	35,68741
68	-0,48953	0,008457	3,206769	-0,139	0,061	0,057	31,38771
<u>69</u>	-0,19892	0,008304	3,321677	-0,245	0,067	0,070	30,71662
70	-0,91594	0,008304	3,393375	-0,205	0,075	0,106	29,27331
<u>71</u>	-0,89008	0,007406	3,429458	-0,382	0,072	0,068	39,33722
72	-0,86315	0,007406	3,405270	-0,265	0,070	0,052	30,86633
73	-0,21443	0,008449	3,077261	0,079	0,054	0,056	28,71841
74	-0,20926	0,010518	3,179944	0,024	0,048	0,064	38,40290
75	-0,44082	0,008397	3,291198	0,194	0,057	0,082	26,67487
76	-0,72899	0,009136	3,387495	-0,111	0,072	0,108	24,47669
77	-0,53933	0,009136	3,302192	-0,209	0,063	0,077	33,90692
78	-0,54586	0,008875	3,309732	-0,277	0,060	0,076	29,89971
79	-0,52518	0,008875	3,220715	-0,140	0,055	0,055	27,37002
80	-0,73525	0,008766	3,395592	-0,178	0,065	0,097	31,94766
81	-0,54178	0,008766	3,324437	-0,137	0,059	0,065	35,95030
82	-0,52028	0,008766	3,257758	-0,258	0,055	0,053	34,81851
83	-0,70178	0,007793	3,445544	-0,346	0,069	0,068	32,78210
84	-0,67511	0,007793	3,424789	-0,373	0,063	0,053	36,02597
<u>85</u>	-0,54913	0,008175	3,225609	-0,151	0,050	0,053	38,10136
86	-0,88600	0,009413	3,387360	-0,130	0,077	0,116	26,58435
87	-0,54586	0,009413	3,301972	-0,197	0,068	0,080	30,28662
88	-0,4811	0,009413	3,20714	-0,145	0,060	0,056	33,26594
89	-0,22341	0,008548	3,308018	-0,209	0,07	0,079	22,4816
90	-0,48327	0,008548	3,216442	-0,094	0,062	0,058	26,60620
91	-0,92546	0,008175	3,395341	0,005	0,074	0,103	25,27724
92	-0,48191	0,008175	3,257674	-0,144	0,061	0,053	32,91770
93	-0,94015	0,007869	3,422750	-0,316	0,069	0,053	32,30415
94	-0,97036	0,007869	3,449755	-0,239	0,085	0,106	28,50492
95	-0,97009	0,007869	3,444121	-0,393	0,076	0,073	28,72995
96	-0,53770	0,009413	3,263564	0,0840	0,063	0,061	24,74549
97	-0,56845	0,008201	3,335263	-0,153	0,075	0,079	27,52030
98	-0,54232	0,008201	3,275810	-0,147	0,069	0,078	27,42417
99	-0,93798	0,007518	3,401738	-0,120	0,087	0,110	24,39900
100	-0,60273	0,007518	3,339844	-0,208	0,078	0,074	29,47733
<u>101</u>	-0,55266	0,008175	3,281677	-0,080	0,065	0,065	30,89188
102	-0,73308	0,007763	3,396672	-0,098	0,068	0,108	22,98455
103	-0,53525	0,007763	3,326401	-0,217	0,059	0,075	31,40947
104	-0,38422	0,007763	3,259726	-0,132	0,054	0,055	28,01910
105	-0,53334	0,007742	3,331333	-0,077	0,058	0,065	40,28682
<u>106</u>	-0,37361	0,007742	3,261746	-0,184	0,055	0,053	36,75212

6.2. Geração dos Modelos de QSAR

Na geração de modelos Quantitativos de Relação Estrutura-Atividade foram utilizadas duas aproximações: Mínimos Quadrados Parciais e Redes Neurais Artificiais. O método PLS baseia-se numa descrição linear da variação da atividade biológica de uma série de compostos em relação a dadas características dos mesmos (descritores) enquanto que as ANN baseiam-se numa abordagem não-linear para resolução do mesmo problema.

6.2.1. Análise de Mínimos Quadrados Parciais (PLS)

A característica mais importante de um modelo QSAR é o seu poder preditivo, que pode ser definido como a habilidade em encontrar uma correlação significativa entre a atividade biológica predita e a observada experimentalmente para compostos que não foram usados na construção do modelo. Isso pode ser feito dividindo-se o conjunto de dados em conjunto de treinamento, composto de 80% dos compostos em estudo, e conjunto de teste, formado pelos 20% restantes, que serão usados para a construção do modelo e para a validação do mesmo, respectivamente.⁴⁹

Os modelos de PLS foram obtidos utilizando o software Pirouette 3.10 com um conjunto de treinamento de 76 compostos, validados externamente à partir de um conjunto teste de 19 compostos. Os estudos de PLS conduzidos com esse conjunto treinamento apontaram para um modelo ótimo com um número de variáveis latentes igual a 4, o qual descreve 86% da variância dos dados. Este modelo foi capaz de prever com alta precisão os valores de p/C_{50} do conjunto teste. A Tabela 4 apresenta os modelos obtidos via PLS bem como os parâmetros estatísticos em que se embasou a escolha do melhor deles.

Tabela 4 - Variância percentual acumulada, SEV, SEC, PRESS, q^2 e r^2 dos modelos PLS obtidos para o conjunto treinamento

VL	Acumulada (%)	SEV	Press Val	q^2	SEC	Press Cal	r^2
1	39,4874	0,349505	9,28369	0,51498	0,340015	8,555134	0,55241
2	56,4198	0,303708	7,01014	0,63400	0,288709	6,084753	0,68166
3	70,2617	0,286262	6,22791	0,67506	0,268923	5,206991	0,72758
4	86,1444	0,276839	5,82462	0,69622	0,259858	4,794373	0,74917
5	91,4254	0,275355	5,76235	0,69964	0,260335	4,744194	0,75179
6	94,3598	0,275716	5,77747	0,69888	0,262135	4,741326	0,75194
7	100,000	0,275634	5,77402	0,69905	0,264051	4,741158	0,75195

A literatura recomenda que um bom modelo de PLS deve apresentar um valor de q^2 acima de 0,5 e de r^2 superior a 0,6. O modelo obtido apresentou valores de q^2 e r^2 iguais a 0,6962 e 0,7492, respectivamente, mostrando que esse é um bom modelo.⁵⁰

Como pode ser visto na Tabela 4, a escolha do modelo com 4 variáveis latentes é justificada por ele apresentar baixos valores de SEV (0,2468) e PRESS (5,825), e o menor valor de SEC (0,2599). Como pode ser notado, os valores para esses erros não diminuem significativamente ao passo que se vai aumentando o número de variáveis latentes dos modelos e, ao mesmo tempo, esse modelo selecionando descreve uma quantidade satisfatória da variância dos dados (86,14%).

A medida que o número de variáveis latentes do modelo vai aumentando acontece um aumento, mesmo que pequeno, nos valores de r^2 e q^2 , já que mais informação vai sendo adicionada ao modelo. No entanto a escolha de um modelo ótimo deve basear-se primeiro nos melhores valores dos erros SEV e SEC. Os menores valores de SEC e SEV foram obtidos nos modelos com 4 e 5 variáveis latentes iguais a 0,2599 e 0,2768 para o primeiro e 0,2603 e 0,2736 para o último. Porém, como o modelo com 5 variáveis latentes não apresentou melhora significativa dos parâmetros estatísticos r^2 e q^2 analisados em relação ao modelo com 4 variáveis latentes (0,71518 e 0,6996 para o primeiro e 0,7492 e 0,6962 para o segundo na ordem em que estão apresentados), o modelo com o menor número de VLs (4) foi escolhido como o melhor modelo.

A contribuição dos descritores para o vetor de regressão do modelo gerado está ilustrada na Figura 11. Nota-se que os descritores que apresentaram os maiores valores de coeficientes de regressão foram HATS1v e RDF110i, sendo portanto os mais relevantes para descrever a atividade biológica do conjunto de compostos estudados. No entanto, a contribuição dos outros descritores não pode ser desprezada uma vez que o conjunto desses descritores apresentam um carácter sinérgico na previsão da atividade biológica e a construção de um modelo preditivo apenas com os descritores em destaque não apresentou valores aceitáveis de q^2 e r^2 .

Figura 11 - Coeficientes de regressão para o modelo PLS.

Para detecção de amostras anômalas com comportamento atípico em relação ao conjunto de compostos estudados, se faz necessário a análise de *outliers*. Que é realizada através da construção do gráfico de alavancagem *versus* resíduos de *Student*. Através da análise da Figura 12 nota-se que a molécula 74 apresenta altos valores de alavancagem embora ainda possua limites aceitáveis de resíduos de student, enquanto que as moléculas 22, 90 e 102 possuem baixa alavancagem, no entanto estão fora dos limites aceitáveis para o resíduo de *Student* (maior que 2 em módulo).

Figura 12 - Detecção de *outliers* no modelo PLS.

Os compostos que se destacaram na análise de outliers tiveram sua estrutura molecular analisada. A Figura 13 ilustra o composto que apresentou alto valor de alavancagem (74) assim também como os compostos que ficaram fora dos limites aceitáveis de resíduo de student (22,90 e 102).

Figura 13 - Compostos identificados na análise de *outliers* com valores altos de alavancagem (composto 74) e valores fora dos limites dos resíduos de Student (compostos 22, 90 e 102). Os substituintes na região R₁ estão evidenciados.

Mesmo que os compostos 22, 74, 90 e 102 tenham apresentado valores indesejados de alavancagem ou resíduos de Student, os modelos gerados sem eles não apresentaram melhora significativa em r^2 ou q^2 por isso o conjunto inicial, incluindo todos os compostos, foi mantido. Observa-se também que as moléculas 74, 90 e 102 apresentaram-se apenas ligeiramente fora dos limites aceitáveis na análise de *outliers* e mesmo o composto 22 não apresenta valores altos de resíduo de Student e alavancagem simultaneamente. É interessante perceber que esses 4 compostos que mostraram comportamento atípico no gráfico de alavancagem *versus* resíduos de Student possuem todos um grupo substituinte volumoso na posição R_1 , o que pode ter contribuído para que essas moléculas se destacassem na análise de *outliers*.

Mesmo após a escolha de um bom modelo, o mesmo deve ser submetido a testes que comprovem a sua robustez, estabilidade e que suas correlações não foram geradas ao acaso. Para tanto, foram realizados testes de validação cruzada com a exclusão de N compostos (*leave-N-out*) e testes de randomização da atividade biológica.

Nos testes de *leave-N-out* foram excluídos de 1 a 16 compostos, sendo realizadas 50 repetições para cada valor de N excluído. De acordo com a literatura para que um dado modelo seja considerado estável ele deve apresentar um valor alto de q^2 e com baixas flutuações a medida que as amostras vão sendo simultaneamente retiradas e valores de desvio para cada valor de N menores que 0,030. Já para que o modelo seja considerado robusto, recomenda-se então que ele apresente valores de q^2 próximos ao correspondente ao teste *leave-one-out* e com desvios inferiores a 0,1. De acordo com os resultados obtidos, o valor de q^2 com a retirada de N amostras se manteve estável (0,6919 em média) e o maior valor de desvio médio registrado foi de 0,015 (ainda assim muito menor que 0,1) com a retirada de 10 amostras, confirmando que o modelo PLS escolhido é estável e robusto.⁵¹

Figura 14 - Testes de *leave-N-out* aplicados ao modelo PLS.

Uma abordagem de validação interna que deve ser usada paralelamente a essas técnicas é a randomização das atividades biológicas. Essa abordagem consiste em repetir o procedimento de construção do modelo com as atividades randomizadas e, em seguida, fazer uma avaliação estatística dos modelos assim produzidos. Se todos os modelos obtidos com as atividades randomizadas apresentarem coeficientes de correlação r^2 e q^2 altos, isso significa que a correlação foi obtida por acaso e não porque a atividade biológica está definitivamente relacionada aos descritores empregados no modelo. Isso implica que não será possível obter um modelo QSAR aceitável para este conjunto de dados utilizando o método que foi empregado.⁵¹

Segundo a literatura, um modelo contruído sem correlações ao acaso deve apresentar valores de q^2 e r^2 menores que 0,4 para modelos randomizados. Como pode ser visto no gráfico da Figura 15, os modelos obtidos na randomização apresentaram valores de r^2 e q^2 menores que 0,2, o que demonstra que o modelo proposto não foi obtido meramente por correlações ao acaso.⁵²

Figura 15 - Teste de randomização aplicado ao modelo PLS.

Ao término dos testes de validação interna anteriormente mencionados, foi realizada então a validação externa por meio da predição dos valores de p/C_{50} do conjunto teste. A predição dos valores de p/C_{50} do conjunto teste apresentou $r^2 = 0,701$, $SEV = 0,413$ e $PRESS = 2,302$.

Como pode ser visto na Figura 16, que confronta os valores de p/C_{50} preditos e experimentais para o conjunto teste, e através da Tabela 5, que apresenta os resíduos de predição, o modelo PLS proposto possui uma boa capacidade preditiva, uma vez que os resíduos são todos menores que 10%, com um valor médio para os resíduos relativos absolutos de 1,78%, sendo o composto 66 o que apresentou o maior resíduo relativo percentual ainda assim igual a apenas 8,26%. O valor dos resíduos relativos dos compostos estudados foi calculado utilizando a expressão (14):

$$\text{Resíduo relativo} = \frac{|\text{Resíduo absoluto}|}{pk_i \text{ experimental}} \cdot 100 \quad (15)$$

Figura 16 - Predição da atividade biológica pelo modelo PLS.

Tabela 5 - Resíduos de predição do conjunto teste no modelo PLS

Composto	pIC ₅₀ Medido	pIC ₅₀ predito	Resíduo absoluto	Resíduo relativo (%)
10	7,987163	7,504354	0,48280	6,0448
11	6,856985	7,163401	0,30642	4,4686
20	7,236572	7,314692	0,07812	1,0795
25	7,124939	6,908393	0,21654	3,0392
27	6,876148	7,064185	0,18804	2,7346
31	7,531652	7,679205	0,14755	1,9590
32	7,856985	7,796352	0,06063	0,7717
34	7,910095	8,221645	0,31155	3,9386
37	6,767004	7,044309	0,27731	4,0979
49	6,928486	7,112871	0,18438	2,6612
52	6,850781	6,972451	0,12167	1,7760
55	7,537602	7,370818	0,16678	2,2126
65	7,327902	7,465757	0,13786	1,8812
66	6,617983	7,164618	0,54664	8,2598
69	6,696804	6,738680	0,04188	0,6253
71	7,244125	7,009086	0,23504	3,2445
85	6,689307	6,677551	0,01175	0,1757
101	6,832683	6,578616	0,25406	3,7184
106	6,531652	6,701439	0,16979	2,5994

Após a realização dos testes de validação interna e externa sobre o modelo PLS proposto pode-se notar que ele mostrou parâmetros estatísticos satisfatórios segundo o recomendado pela literatura. Como pode ser visto, a análise do gráfico de alavancagem *versus* resíduos de Student não detectou nenhuma amostra atípica, assim também os testes de *leave-N-out* e randomização confirmaram tanto a robustez do modelo quando a inexistência de correlações ao acaso. O modelo proposto também teve sua capacidade preditiva comprovada por meio da validação externa realizada a partir da predição dos valores de pIC_{50} do conjunto teste. Desde modo, pode-se afirmar que o modelo PLS obtido apresenta capacidade preditiva comprovada e pode servir como uma ferramenta útil na busca por novos compostos com alta atividade biológica frente aos receptores 5-HT_{2a}.

6.2.2. Análise de ANN

A criação de um modelo preditivo por meio da técnica ANN permite avaliar a existência de uma relação não-linear entre os descritores e a atividade biológica. A fim de investigar se os descritores selecionados e a atividade biológica apresentavam uma relação não-linear uma rede neural do tipo MLP-ANN foi utilizada para prever os valores de pIC_{50} do conjunto de compostos arilpiperazínicos estudados em relação ao receptor 5-HT_{2a}.

Para esses testes foram escolhidas funções de transferência *tansig-tansig* dado que vários outros trabalhos anteriores realizados neste grupo mostraram que esse tipo de função apresenta bons resultados na geração de modelos preditivos de fármacos.^{53,54,55} Inicialmente foram usados 7 neurônios (os sete descritores selecionados) na camada de entrada e foram realizados vários testes para se definir o número ideal de neurônios na camada intermediária. Nesta etapa variou-se o número de neurônios na camada intermediária de 1 a 20.

Para a função de transferência utilizada, determinou-se o número de neurônios na camada intermediária capaz de produzir o modelo com os melhores valores de EQM e DMA. Além disso, avaliou-se também os valores de r^2 para os conjuntos de treinamento, teste e validação utilizados no

treinamento da MLP-ANN. A Tabela 6 mostra o desempenho da MLP-ANN treinada através dos valores de erro quadrático médio (EQM) dos melhores modelos obtidos no decorrer da variação do número de neurônios da camada interna.

Tabela 6 – Melhores modelos de MLP-ANN gerados para cada camada intermediária

Número de Neurônios	EQM	DMA	$r^2_{\text{treinamento}}$	$r^2_{\text{validação}}$	r^2_{teste}
1	0,011705	0,087443	0,619293	0,918975	0,899121
2	0,009864	0,080792	0,744982	0,940716	0,764261
3	0,009842	0,079633	0,749488	0,941276	0,729892
4	0,009854	0,082166	0,791159	0,880222	0,635088
5	0,009983	0,08179	0,74728	0,885238	0,757728
6	0,009945	0,078757	0,765088	0,898151	0,685342
7	0,009880	0,080935	0,756003	0,947628	0,796696
8	0,009428	0,078224	0,742051	0,724576	0,920684
9	0,009532	0,075909	0,79380	0,879011	0,683119
10	0,009637	0,077526	0,793755	0,794587	0,787618
11	0,009877	0,079908	0,780543	0,905939	0,720399
12	0,009648	0,077393	0,76013	0,710981	0,892793
13	0,009957	0,079103	0,791948	0,869374	0,546395
14	0,00995	0,078151	0,743791	0,802865	0,855363
15	0,009943	0,078548	0,796939	0,774741	0,663943
16	0,009896	0,076698	0,787172	0,784075	0,694925
17	0,009885	0,080276	0,660344	0,820697	0,835288
18	0,011574	0,083925	0,751095	0,871663	0,702651
19	0,009586	0,077487	0,791744	0,865581	0,733850
20	0,009684	0,080223	0,788193	0,695122	0,806011

O treinamento das MLP-ANN com funções de transferência tansig-tansig mostrou que 10 era o número ótimo de neurônios na camada intermediária da rede neural por apresentar valores baixos de Erro Quadrático Médio e Desvio Médio Absoluto (EQM =0,894 e DMA = 0354) e $r^2_{\text{validação}}$, $r^2_{\text{treinamento}}$ de r^2_{teste} próximos entre si sendo o valor de r^2_{teste} o menor dentre esses três, o que é um indício de que o modelo escolhido não apresenta super ajuste da rede. Deste modo a arquitetura da MLP-ANN selecionada foi 7-10-1 (sete neurônios na camada de entrada, 10 neurônios na camada intermediária e 1 neurônio na

camada de saída, como ilustrado na Figura 17) sendo que a rede escolhida foi treinada até a interação 5.172.

Figura 17 – Representação da MLP-ANN selecionada.

No método MLP-ANN as etapas de validação e calibração são realizadas simultaneamente ao treinamento da rede, de onde se obtêm os valores de r^2 do conjunto de treinamento, validação e teste. Para o modelo selecionado os valores dessas variáveis foram iguais a 0,794587, 0,793755 e 0,787618, respectivamente. Tal modelo também apresentou baixos valores de erros de predição da atividade biológica do conjunto teste, como ilustrado pela Figura 18 e apresentado na Tabela 7.

Figura 18 - Predição da atividade biológica pelo modelo ANN.

Tabela 7 – Resíduos de predição do conjunto teste pelo modelo ANN

Composto	pIC_{50} experimental	pIC_{50} predito	Resíduo absoluto	Resíduo relativo (%)
11	6,856985	5,995249	6,856985	5,520940
12	7,698970	5,995249	7,698970	3,283879
20	7,236572	5,995249	7,236572	4,200129
25	7,124939	5,995249	7,124939	3,860729
27	6,876148	5,995249	6,876148	2,818466
31	7,531653	5,995249	7,531653	3,696555
32	7,856985	5,995249	7,856985	0,120148
34	7,910095	5,995249	7,910095	0,299257
37	6,767004	5,995249	6,767004	4,130149
49	6,928486	5,995249	6,928486	3,507631
52	6,850781	5,995249	6,850781	1,059406
55	7,537602	5,995249	7,537602	1,033389
65	7,327902	5,995249	7,327902	1,446216
66	6,617983	5,995249	6,617983	5,880572
69	6,696804	5,995249	6,696804	0,362340
71	7,244125	5,995249	7,244125	0,179268
85	6,689307	5,995249	6,689307	2,951926
101	6,832683	5,995249	6,832683	3,185785
106	6,531653	5,995249	6,531653	0,284043

Como pode ser visto os valores de pIC_{50} preditos para o conjunto teste apresentam erros relativos menores que 10%. Nota-se que o composto que apresentou o maior desvio percentual relativo foi o composto 66, com o valor de 5,88%. De um modo geral os valores de pIC_{50} preditos pelo modelo de ANN foi muito próximo do valor experimental dessa medida, apresentando um resíduo relativo percentual médio de apenas 2,95%. Assim, pode-se afirmar que o modelo ANN apresenta boa capacidade preditiva dos valores de pIC_{50} do conjunto de compostos em estudo, o que mostra que a atividade biológica dessas arilpiperazinas frente ao receptor 5-HT_{2a} também é bem descrita através de uma aproximação não-linear.

6.3. Análise dos Descritores Seleccionados

Os descritores utilizados na construção dos modelos de ANN e de PLS foram seleccionados via Algoritmo Genético e estão apresentados de maneira resumida pela Tabela 8. Os descritores codificam a informação química implícita nas moléculas de modo que elas possam ser utilizadas na construção de modelos matemáticos com capacidade preditiva. Um bom entendimento do significado químico de um descritor pode ajudar no planejamento racional de compostos com alta atividade biológica.

Tabela 8 – Descritores seleccionados e suas definições

Descritor	Tipo	Propriedade Relacionada	Definição
E_{LUMO2+}	Eletrônico	Energia do LUMO	Terceiro Orbital Desocupado de Menor Energia
HATS1v	GETAWAY	Volume de substituintes	Alavancagem de autocorrelação1/ balanceada por volumes atômicos de van der Waals
R3p+	GETAWAY	Polarizabilidade	Autocorrelação máxima de ordem 3/ balanceada por polarizabilidades
RDF110i	Molecular	Potencial de Ionização	Distribuição radial balanceada por potenciais de ionização
SpMax6_Bhm	Molecular	Momento de Dipolo	Maior auto valor de uma matriz de espectros
Mor27p	3D-MoRSE	Polarizabilidade	Descritor 3D-MoRSE balanceado por polarizabilidades
JGI9	Topológico	Transferência de carga	Índice de carga topológica média de ordem 9

6.3.1. E_{LUMO2+}

O orbital ocupado de mais alta energia HOMO (do inglês *highest occupied molecular orbital*) é o orbital molecular que, de acordo com o princípio da construção, é ocupado por último e o orbital desocupado de mais baixa energia LUMO (do inglês *lowest unoccupied molecular orbital*) é, então, o próximo orbital molecular. Estes orbitais são também conhecidos como orbitais de fronteira e são de grande importância para a interpretação de estruturas e em estudos cinéticos.⁵⁶

Uma lição que tem sido aprendida nos últimos anos é que os orbitais LUMO não são meros espectadores, eles comumente desempenham um papel chave na determinação do resultado de uma reação química (isso significa que eles estão de fato envolvidos na ligação). Portanto não é surpresa que exista evidência tanto teórica quanto experimental que a doação de elétrons entre compostos químicos envolva a presença de orbitais LUMO.⁵⁷

No caso do descritor selecionado, E_{LUMO2+} , tem-se os valores de energia do terceiro orbital desocupado de cada composto. Nota-se que baixos valores de energia do LUMO aumentam a probabilidade do composto em aceitar elétrons.

Figura 19 - Compostos que apresentaram baixos valores de E_{LUMO2+} .

Ao analisar os dados do descritor $E_{\text{LUMO}2+}$ para os compostos arilpiperazínicos sob estudo, percebe-se que, os compostos que apresentam maior atividade biológica, com valores de p/C_{50} em torno de 8 (Figura 19), apresentaram valores $E_{\text{LUMO}2+}$ próximos de 7eV. A observação de o valor da energia para o descritor LUMO 2+ nessas moléculas encontrar-se próximo de 7eV, sugere que esta deve ser uma energia ótima para que a interação entre tais compostos e os sítios ativos do receptor 5-HT_{2a} seja favorecida. Observa-se também que a região em que os orbitais LUMO 2+ se concentram nas moléculas de mais alta atividade hora se encontra sobre o anel piperazínico, hora no anel de benzeno ligado a porção imidazol dessas moléculas, evidenciando que essas duas regiões devem ser de grande importância para o estabelecimento da interação entre o composto teste e o receptor 5-HT_{2a}. Espera-se que em estudos posteriores essas duas hipóteses possam ser investigadas a fundo com a realização de estudos de *docking* e dinâmica molecular através da simulação da interação desses compostos com o sítio ativo do receptor em questão.

6.3.2. HATS1v

O descritor HATS1v, ou alavancagem de autocorrelação 1/ balanceada por volumes atômicos de van der Waals, é um descritor topológico do tipo GETAWAY. Os descritores *GETAWAY* (do inglês *Geometry, Topology and Atom-Weights Assembly*) são descritores geométricos que codificam a informação da posição efetiva de substituintes e fragmentos no espaço molecular. Eles são, no entanto, independentes do alinhamento da molécula e até certo grau também levam em conta informações a respeito do tamanho, formato e de algumas propriedades atômicas específicas.

Vale ressaltar que compostos com grupos substituintes volumosos nas posições R₁ e R₂ apresentaram altos valores para este descritor, como os compostos **7**, **94** e **99** com valores de 0,080, 0,085 e 0,087, respectivamente. A Figura 20 apresenta algumas das moléculas com altos valores para o descritor HATS1v.

Figura 20 – Compostos que apresentaram altos valores para o descritor HATS1v. Em vermelho estão destacados os substituintes volumoso nas posições R₁ e R₂ das moléculas.

Os descritores *GETAWAY* foram propostos recentemente como descritores químico-estruturais oriundos de uma nova representação da estrutura molecular, a Matriz de Influência Molecular (MIM), representada por H e definida como:

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T \quad (16)$$

onde M é a matriz molecular consistindo nas coordenadas Cartesianas centradas x, y, z dos átomos da molécula (hidrogênios incluídos) em uma conformação escolhida, e o superescrito T se refere à matrix transposta. Assume-se que as coordenadas atômicas são calculadas em relação ao centro geométrico da molécula para se obter assim a invariância translacional.

A equação precisa que descreve o descritor HATS está apresentada abaixo⁵⁸:

$$HATS_k(w) = \sum_{i=1}^{A-1} \sum_{j>i} (w_i \cdot h_{ii}) \cdot (w_j \cdot h_{jj}) \cdot \delta(k; d_{ij}) \quad k = 0, 1, 2, \dots, d \quad (17)$$

onde A é o número de átomos na molécula;

d é o diâmetro topológico;

d_{ij} é a distância topológica entre os átomos i e j;

e w_i é o peso físico-químico do iésimo átomo;

e $\delta(k; d_{ij})$ é uma função delta de Dirac definida como:

$$\delta(k; d_{ij}) = \begin{cases} 1 & \text{se } d_{ij} = k \\ 0 & \text{se } d_{ij} \neq k \end{cases} \quad (18)$$

6.3.3. R3p+

O descritor *GETAWAY* R3p+ (autocorrelação máxima de ordem 3/ balanceada por polarizabilidades) está relacionado à geometria molecular tridimensional, relação de átomos e informação química calculadas através de fórmulas de autocorrelações espaciais, índices de conectividade e algumas operações matriciais. Este descritor leva em conta os átomos na molécula que influenciam na polarizabilidade, ou seja, o descritor R3p+ está relacionado ao formato da molécula, tamanho e polarizabilidades atômicas e contem informações a respeito da presença de substituintes marcantes nas proximidades dos átomos terminais.^{59,60}

Nota-se que os compostos que apresentaram valores expressivos para o descritor R3p+ possuem átomos altamente polarizáveis nos substituintes R1, R4 e R5, em especial átomos de cloro e oxigênio nas extremidades das moléculas, como pode ser visto nos compostos **18**, **76** e **102** com valores de R3p+ iguais a 0,116, 0,108 e 0,108 na ordem em que estão apresentados. A Figura 21 ilustra os átomos altamente polarizáveis presentes nos substituintes R1, R4 e R5 das moléculas **18**, **76** e **102**.

Figura 21 - Moléculas que apresentaram valores altos para o descritor R3p+. Destacado em vermelho estão átomos com alta polarizabilidade presentes nas posições R1, R4 e R5 destas moléculas.

6.3.4. RDF110i

O descritor RDF110i consiste na função de distribuição radial balanceada por potenciais de ionização. A função de distribuição radial (RDF) de um conjunto de N átomos pode ser interpretada como a distribuição de probabilidades de se encontrar um átomo em um volume esférico de raio r. A equação que representa a função de distribuição radial é:

$$g(r) = f \sum_i^{N-1} \sum_{j>i}^N A_i A_j e^{-B(r-r_{ij})^2} \quad (19)$$

onde f é um fator de escalonamento;

N é o número de átomos;

e A são propriedades características dos átomos i e j.

Nessa equação o termo exponencial contém a distância r_{ij} entre os átomos i e j e o parâmetro de suavização B , que define a distribuição da probabilidade das distâncias individuais. As propriedades atômicas A_i e A_j , usadas na equação, permitem a discriminação dos átomos de uma molécula para quase todas as propriedades que podem ser atribuídas a ele.⁶¹

Como pode ser visto na Figura 22, compostos com altos valores para este descritor apresentaram baixas energias dos orbitais HOMO, em comparação ao grupo de moléculas estudadas, propriedade essa intimamente relacionada à energia de ionização, uma vez que moléculas com baixos valores de HOMO são mais facilmente ionizadas.

Figura 22 – Moléculas que apresentaram altos valores para o descritor RDF110i.

6.3.5. SpMax6_Bhm

SpMax é o autovalor principal, isto é, o maior auto valor de uma matriz de espectros. Este descritor é balanceado pelo momento de dipolo. O momento de dipolo por sua vez está relacionado a polaridade da molécula. O surgimento do momento de dipolo é devido a assimetria na estrutura tridimensional e distribuição de cargas de uma dada substância química. Pode-se notar que moléculas com uma distribuição assimétrica de carga devido a presença de átomos bastante eletronegativos em seus substituintes apresentaram valores acentuados para este descritor. Na Figura 23 estão ilustradas as moléculas **43**, **53** e **94**, que apresentam valores expressivos para este descritor iguais a 3,43, 3,42 e 3,45, respectivamente. Vale ressaltar que essas moléculas possuem regiões com elevados momento de dipolo devido a presença de ligações de átomos de carbono com átomos de alta eletronegatividade como oxigênio, flúor e cloro como ressaltado na figura.

Figura 23 – Moléculas que apresentaram altos valores para o descritor SpMax6_Bhm. A presença de unidades assimétricas de carga devido à átomos bastante eletronegativos nas extremidades das moléculas está evidenciada na cor vermelha.

Índices espectrais são descritores moleculares definidos em termos de autovalores da matrix quadrada M de tamanho $(n \times n)$. Os autovalores são as raízes do polinômio característico da matriz M e o conjunto de autovalores é a matriz de espectro $A(M) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$; os autovalores são convenientemente nomeados de modo que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. A função de autovalores SpMax utilizada na obtenção de índices espectrais, pode ser aplicada a quaisquer matrizes moleculares $M(w)$, calculadas com o artifício de balanceamento w .⁶² A representação matemática de um descriptor SpMax pode ser escrita como:

$$\text{SpMax} (M,w) = \max_i \{\lambda_i\} \quad (20)$$

6.3.6. Mor27p

Descritores 3D-MoRSE (do inglês *Molecular Representation of Structures based on Electronic diffraction*) foram apresentados em 1996 por J.H. Schuur, P. Selzer e J.Gasteiger com a motivação de codificar a estrutura tri-dimensional de uma molécula a partir de uma quantidade fixa de variáveis. Mor27p é um descriptor do tipo 3D-MoRSE balanceado por polarizabilidades. Para calcular representações tri-dimensionais de estruturas baseadas em descritores de difração de elétrons (3D-MoRSE) Gasteiger et al, retornou à curva $I(s)$ inicial e manteve a sua forma explícita. Para o artifício de peso atômico w , várias propriedades fisico-químicas tais como massa atômica, cargas atômicas parciais, polarizabilidade atômica são consideradas. Com o propósito de se obter descritores de comprimento uniforme, é imposto que a distribuição de intensidades $I(s)$ (Equação 20) seja discreta, com seu valor calculado como uma sequência de valores bem distribuídos, por exemplo, 32 ou 64 valores de $1 - 31 \text{ \AA}^{-1}$.

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}} \quad (21)$$

onde s é o parâmetro de espalhamento;

r_{ij} é a distância euclidiana entre os átomos i e j ;

N é o número total de átomos;

e A_i e A_j são diferentes propriedades atômicas usadas como pesos.⁶³

Como pode ser visto na Figura 24 compostos como átomos altamente polarizáveis apresentaram valores apreciáveis para este descritor, como as moléculas **14**, **34** e **53** iguais a -0,44, -0,46 e -0,50, respectivamente. Vale a pena ressaltar aqui que o composto **34**, é o único dentre os derivados de arilpiperazinas estudados que possui um grupo substituinte triazol.

Figura 24 – Compostos que apresentaram valores apreciáveis para o descritor Mor27p. Em vermelho estão evidenciados os átomos com alta polarizabilidade presentes nas extremidades dos grupos substituintes dessas moléculas.

6.3.7. JGI9

JGI9 significa índice de carga topológica média de ordem 9. Índices de carga topológica são propostos para avaliar a transferência de carga entre pares de átomos e está intimamente relacionado as ramificações da molécula.

Para obter os índices de carga primeiro defini-se a matriz M como:

$$\mathbf{M} = \mathbf{A} \cdot \mathbf{D}^{-2} \quad (22)$$

onde A é a matriz adjacência;

$\mathbf{D}^{-2}$ é o recíproco quadrado da matriz de distâncias;

e M é a matriz de Galvez $A \times A$ em que A é o número de átomos na molécula.

A matriz M dá origem a uma matriz assimétrica de carga, CT definida como:

$$CT_{ij} = \begin{cases} \delta_i & \text{se } i = j \\ m_{ij} - m_{ji} & \text{se } i \neq j \end{cases} \quad (23)$$

onde os m_{ij} são elementos de M e δ_i é a posição do vértice de i.

Os valores de entrada diagonais de CT representam a valência topológica dos átomos e as entradas não diagonais e CT_{ij} representam a quantidade de carga transferida de um átomo J para o átomo i.

Para cada caminho de comprimento K pode-se definir o índice de carga topológico G_k como:

$$g_k = \frac{1}{2} \cdot \sum_{i=1}^A \sum_{j=1}^A |CT_{ij}| \cdot \delta(k; d_{ij}) \quad (24)$$

onde $\delta(k; d_{ij})$ é o delta de Kronecker,

$$\delta(k; d_{ij}) = \begin{cases} 1 & \text{se } d_{ij} = k \\ 0 & \text{se } d_{ij} \neq k \end{cases} \quad (25)$$

em que d_{ij} são os elementos da matriz de distâncias.

Assim, G_k representa a transferência de carga total entre os átomos em uma distância topológica k . O valor máximo dos termos G_k é igual ao diâmetro topológico, D .

O índice de carga topológica média, J_k é definido como:

$$J_k = \frac{G_k}{A - 1} \quad (26)$$

e o índice de carga topológica global, J , definido como:⁶⁴

$$J = \sum_{k=1}^5 J_k \quad (27)$$

Foi possível notar que compostos que apresentavam átomos de alta eletronegatividade, como oxigênio, flúor e cloro diretamente ligados a átomos de carbono nas extremidades dos substituintes R_1 , R_4 e R_5 e substituintes metila nas porções R_2 e R_3 alcançaram altos valores para o descritor JGI9. A presença dessas ligações químicas altamente polares nas moléculas dá origem a uma transferência de carga entre átomos vizinhos, que é o fenômeno marcante característico deste descritor. Na Figura 25 estão representados os compostos **76**, **77** e **86** que apresentaram valores iguais a 0,0091, 0,0091 e 0,0094 para o descritor JGI9.

Figura 25 – Compostos que apresentaram valores apreciáveis para o descritor JGI9. Em vermelho estão destacadas as regiões de transferência de carga entre átomos vizinhos.

7. CONCLUSÕES E PERSPECTIVAS

O estudo de QSAR dos compostos arilpiperazínicos investigados utilizando os métodos PLS e ANN permitiu propor dois modelos matemáticos distintos, capazes de prever a atividade biológica dessa classe de compostos em relação ao receptor 5-HT_{2a}. Características importantes das estruturas dos compostos foram identificadas a partir dos descritores selecionados. Observou-se que, de forma geral, os compostos mais ativos apresentaram altos valores dos descritores HATS1v, SpMax6_Bhm, JGI9, RDF110i e baixos valores para os descritores E_{Lumo2+} e Mor27p.

Os modelos matemáticos gerados via PLS apresentaram parâmetros estatísticos nos limites aceitáveis indicados pela literatura. Os testes de validação realizados comprovaram sua capacidade preditiva, robustez e estabilidade e demonstraram que os mesmos não foram obtidos a partir de correlações ao acaso. Já as análises de MLP-ANN permitiram identificar a arquitetura 7-10-1, treinada com as funções de transferência *tansig-tansig*, como a mais adequada para descrição da atividade biológica dos compostos investigados. A análise dos erros EQM e DMA, assim como dos parâmetros r^2 e q^2 , comprovou a estabilidade estatística do modelo obtido via redes neurais.

A proposição de dois modelos de QSAR distintos, um via PLS e outro via ANN, permitiu observar que tanto relações lineares quanto não lineares são capazes de descrever de forma satisfatória as relações entre o conjunto de descritores selecionado e a atividade biológica dos compostos em estudo. Portanto, pode-se afirmar que os dois modelos de QSAR aqui propostos mostram-se promissores para a predição da atividade biológica de novos compostos da classe arilpiperazínicos frente ao receptor 5-HT_{2a}, o que pode vir a contribuir significativamente no planejamento de novos fármacos mais eficazes no tratamento do distúrbio depressivo maior.

Dando continuidade a este trabalho, visa-se a realização de estudos de *docking* e simulações por dinâmica molecular a fim de se alcançar um melhor entendimento da interação e do mecanismo de ação dos compostos arilpiperazínicos estudados em relação ao sítio ativo do receptor 5-HT_{2a}.

8. REFERÊNCIAS

1. Depression. NATIONAL INSTITUTE OF MENTAL HEALTH- NIH. Disponível em: <<https://www.nimh.nih.gov/health/topics/depression/index.shtml>>. Acesso em 10 nov. 2016
2. Depression. WORLD HEALTH ORGANIZATION- WHO. Disponível em: <<http://www.who.int/mediacentre/news/releases/2016/depression-anxiety-treatment/en/>>. Acesso em 10 nov. 2016).
3. CONRADI, H. J., ORMEL, J., de JONGE, P. Depression: A Global Crisis. Disponível em: <http://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf>. Acesso em 14 nov. 2016
4. Cases of depressive disorder. UN WORLD HEALTH ORGANIZATION- WHO- WHO. Disponível em: <<http://www.un.org/apps/news/story.asp?NewsID=56230#.WcBtDSMrJFZ>>. Acesso em: 13 nov. 2016.
5. Depressão já é a doença mais incapacitante, afirma OMS. FOLHA DE SÃO PAULO. Disponível em: <<http://www1.folha.uol.com.br/equilibrioesaude/2014/12/1563458-depressao-ja-e-a-doenca-mais-incapacitante-afirma-a-oms.shtml>> Acesso em 10 de out. 2015.
6. ROMEIRO, L. A. S.; FRAGA, C. A. M.; BARREIRO, E. J. Novas estratégias terapêuticas para o tratamento da depressão: uma visão da química medicinal. **Química Nova**, v.26, p.347-358, 2003.
7. SEO, H.J.; PARK, E.J.; KIM, M. J.; KANG, S. Y.; LEE, S. H.; KIM, H. J.; LEE, K. N.; JUNG, M. E.; LEE, M. E.; KIM, M. S.; SON, E. J.; PARK, W. K.; KIM, J.; LEE, J. Design and Synthesis of Novel Arylpiperazine Derivatives Containing the Imidazole Core Targeting 5-HT_{2a} Receptor and 5-HT transporter. **Journal of Medicinal Chemistry**. v. 54, p. 6305-6318, 2011.
8. JOSHU, K.; DELORENZO, C.; CHOUDHURY, S.; PARSEY, R.V. The 5-HT_{1a} receptor in Major Depressive Disorder. **European Neuropsychopharmacology**, v. 26, p.397-410 2015.
9. BULTINCK, P.; WINTER, H.; LANGENEAKER, W.; TOLLENARE, J. P. Computational Medicinal Chemistry for Drug Discovery. **CRC Press**, ed. 1, 2003. 1169p.
10. HOLDEN, C. Future Brightening for Depression Treatments. *Science*, v.302, p.810-813, 2003.

11. PIERCE, K. L.; PREMONT, R. T.; LEFKOWITZ, R. J. Seven-transmembrane receptors. **Nature Reviews Molecular Cell Biology**, v. 3, p. 639-650, 2002.
12. RUBENSTEIN, L. A.; LANZARA, R. G. Activation of G protein-coupled receptors entails cysteine modulation of agonist binding. **Journal of Molecular Structure (Theochem)**, v.430: p. 57–71, 1998.
13. JENCK, F.; MOREAU, J. L.; MUTEL, V.; MARTIN, J. R. 5HT_{2C} receptor agonists exhibit antidepressant-like properties in the anhedonia model of depression in rats. **European Neuropsychopharmacology**, v. 6, n. 6, p.169-175, 1996.
14. SHARP, T.; UMBERS, V.; GARTSIDE, S. Effect of a selective 5-HT reuptake inhibitor in combination with 5-HT_{1A} and 5-HT_{1B} receptor antagonists on extracellular 5-HT in rat frontal cortex in vivo. **British Journal of Pharmacology**, v.121, p.941-946, 1997.
15. LEVITT, P.; PINTAR, J. E.; BREAKFIELD, X. O. Immunocytochemical demonstration of monoamine oxidase B in brain astrocytes and serotonergic neurons. **Proc Natl Acad Sci U S A**, v. 79, n.20, p. 6385-9, 1982.
16. BALDESSARINI, R. J. Current status of antidepressants: Clinical pharmacology and therapy. **The Journal of Clinical Psychiatry**, v. 50, n.4, p. 117-126, 1989.
17. MURPHY, D. L.; WICHEMS, C. L.; HEILS, A. Molecular manipulations as tools for enhancing our understanding of 5-HT neurotransmission. **Trends in Pharmacological Science**, v. 20, p. 246-252, 1999.
18. SOMMI, R. W.; CRISMON, M. L.; BOWDEN, C. L. Fluoxetine: A Serotonin-specific, Second-generation Antidepressant. **Pharmacotherapy**, v. 7, n. 1, p. 1-14, 1987.
19. HOYER, D., HANNON, J. P., AND MARTIN, G. R. Molecular, pharmacological and functional diversity of 5-HT receptors. **Pharmacology Biochemistry and behavior**, v. 71, p. 533-554, 2002.
20. GOODARZI, M.; TIMOFEI, S. F.; HEYDEN, Y. V. Towards better understanding of feature-selection or reduction techniques for Quantitative Structure–Activity Relationship Models. **Trends in Analytical Chemistry**, v. 42, 2013.
21. WINKLER, D. A.; MOMBELLIC, E.; PIETROIUSTI, A.; TRANE, L.; WORTHF, A.; FADEELG, B.; McCLLH, M. J. Applying quantitative structure–activity relationship approaches to nanotoxicology: Current status and future potential. **Toxicology**, v. 313, p.15-23, 2013.
22. KUBINYI, H. QSAR: Hansch analysis and related approaches. New York: **VCH**, 1993. 240p.

- 23.ASIRVATHAM, S.; DHOKCHAWLE, B.V.; TAURO, S. J.; Quantitative structure activity relationships studies of non-steroidal anti-inflammatory drugs: A review. **Arabian Journal of Chemistry**, in press, 2016.
- 24.PERKINS, R.; FANG, H.; TONG, W.; WELSH, W.J. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. **Environmental Toxicology and Chemistry**, v. 22, n. 8, p. 1666–1679, 2003.
- 25.VAN DE WATERBEEMD, H.; ROSE, S. Quantitative approaches to structure-activity relationships. In: WERMUTH, C. G. (Ed.). The practice of medicinal chemistry. 2.ed. London: **Elsevier Academic Press**, 2003.
- 26.FOYE, W. O.; LEMKE, T. L.; WILLIAMS, D. A. Principles of medicinal chemistry. 4.ed. **Baltimore: Williams & Wilkins**, 1995. 995p.
- 27.BROWN, N.; LEWIS, R. A. Exploiting QSAR methods in lead optimization. **Current Opinion in Drug Discovery**, v.9, p.419-424, 2006.
- 28.SABIN, J. R.; TRICKEY, S. B.; APELL, S. P.; ODDERSHEDE, J. Molecular Shape, Capacitance, and Chemical Hardness. **International Journal of Quantum Chemistry**, v. 77, p. 358–366, 2000.
- 29.PASHA, F.A., SRIVASTAVA, H.K., SINGH, P.P. Comparative QSAR study of phenol derivatives with the help of density functional theory. **Bioorganic & Medicinal Chemistry**, v. 13, p. 6823–6829, 2005.
- 30.KO, G. M.; REDDY, A. S.; GARG, R.; KUMAR, S.; HADAEGH, A. R. Computational Modeling Methods for QSAR Studies on HIV-1 Integrase Inhibitors. **Current Computer-Aided Drug Design**, v. 8, p. 255-270, 2012.
- 31.FERREIRA, M. M. C.; MONTANARI, C. A.; GAUDIO, A. C. Seleção de Variáveis em QSAR. **Química Nova**, v. 25, n. 3, p. 439-448, 2002
- 32.SHAHLAEI, M.; Descriptor Selection Methods in Quantitative Structure–Activity Relationship Studies: A Review Study. **Chem. Rev.** V. 113, p. 8093–8103, 2013.
- 33.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**, v. 20, n. 4, p. 770–787, 2009.
- 34.GAUDIO, A.C.; ZADONA de, E., Preposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. **Química Nova**, n. 5, v. 24, 2001.
- 35.ATA, R. Artificial neural networks applications in wind energy systems: a review. **Renewable and Sustainable Energy Reviews**. V.49, P.534–562, 2015.

36.GARDNER, M. W.; DORLING, S. R. Artificial Neural Networks (The Multilayer Perceptron) – A Review of Applications in the Atmospheric Sciences. **Atmospheric Environment**, v. 32, n. 14/15, p. 2627-2636, 1998.

37.HAYKIN, S. Neural networks: a comprehensive foundation. 2. ed. Upper Saddle River:**Prentice Hall**, 1999. 842 p.

38.FERREIRA, M. M. C.; KIRALJ, R. Métodos quimiométricos em relações quantitativas estrutura-atividade (QSAR). In: MONTANARI, C. A. (Ed.). Química medicinal: métodos e fundamentos em planejamento de fármacos. São Paulo: **Editora da Universidade de São Paulo**, 2011. p. 387-453.

39.TROPSHA, A. Application of predictive QSAR Models to Database Mining. In OPREA, T. (Ed.). Chemoinformatics in Drug Discovery, **Weinheim: Wiley-VCH**, 2005.

40.GAUSSIAN. FRISCH, M. J., TRUCKS, G. W., SCHLEGEL, H. B., SCUSERIA, G. E., ROBB, M. A., CHEESEMAN, J. R., SCALMANI, G., BARONE, V., MENNUCCI, B., PETERSSON, G. A., NAKATSUJI, H., CARICATO, M., LI, X. HRATCHIAN, H.P., IZMAYLOV, A. F., BLOINO, J., ZHENG, G., SONNENBERG, J. L.; HADA, M., EHARA, M., TOYOTA, K., FUKUDA, R., HASEGAWA, J., ISHIDA, M., NAKAJIMA, T., HONDA, Y., KITAO, O., NAKAI, H., VREVEN, T., MONTGOMERY, JR., J. A., PERALTA, J. E., OGLIARO, F., BEARPARK, M., HEYD, J. J., BROTHERS, E., KUDIN, K N., STAROVEROV, V. N., KOBAYASHI, R.,NORMAND, J., RAGHAVACHARI, K., RENDELL, A., BURANT, J. C., IYENGAR, S.S., TOMASI, J., COSSI, M., REGA, N., MILLAM, N. J., KLENE, M.; KNOX, J. E., CROSS, J. B., BAKKEN, V., ADAMO, C., JARAMILLO, J., GOMPERS, R., STRATMANN, R. E., YAZYEV, O., AUSTIN, A. J., CAMMI, R., POMELLI, C., OCHTERSKI, J. W., MARTIN, R. L., MOROKUMA, K., ZAKRZEWSKI, V. G., VOTH, G. A., SALVADOR, P., DANNENBERG, J. J., DAPPRICH,S., DANIELS, A.D., FARKAS,O., FORESMAN, J. B., ORTIZ, J. V., CIOSLOWSKI, J., FOX, D. J. Gaussian, Inc., Wallingford CT, 2009.

41.LEE, C.; YANG,W.; PARR, R. G.; Physical Review B, n. 37, v. 785, 1988.

42.POOPLE, J. A.; SEEGER, R.; BINKLEY, J. S.; KRISHNAN, R. Self-Consistent Molecular Orbital Methods. XX. A Basis Set for Correlated Wavefunctions. Journal of Chemical Physics, v. 72, 650, 1980.GOLBRAIKH, A.; TROPSHA, A. Beware of q^2 ! **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 269–276, 2002

43.TODESCHINI, R.; CONSONNI, V.; PAVAM, M. Dragon 2.1. Milano, 2002.

44.Yap CW (2011). PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry. 32 (7): 1466-1474.

45. OLIVEIRA, D. B.; GAUDIO, A. C. BuildQSAR : A new computer program for QSAR analysis. **Quantitative Structure-Activity Relationships**, v. 19, n. 6, p. 599–601, 2001.
46. INFOMETRIX INC. Pirouette 3.11. Woodinville, 2002.
47. MARTINS, J. P. A.; FERREIRA, M. M. C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. **Química Nova**, v. 36, n. 4, p. 554–560, 2013.
48. MATHWORKS. Matlab 7.12. 2011.
49. TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. **QSAR & Combinatorial Science**, v. 22, n. 1, p. 69–77, 2003.
50. WOLD, S.; ERIKSSON, L.; CLEMENTI, S. Statistical validation of QSAR results. In: WATERBEEMD, H. VAN DE (Ed.). Chemometric methods in molecular design. New York: **VCH**, 1995. p. 309–338.
51. GRAMATICA, P. Principles of QSAR models validation: internal and external. **QSAR and Combinatorial Science**, v. 26, p. 694-701, 2007.
52. ERIKSSON, L.; JAWORSKA, J.; WORTH, A. P.; CRONIN, M. T. D.; MCDOWELL, R. M.; GRAMATICA, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. **Environmental Health Perspectives**, v. 111, n. 10, p. 1361–1375, 2003.
53. OLIVEIRA, A.A. QSAR e dinâmica molecular no estudo de sistemas biomoleculares: predição da atividade biológica de antagonistas do receptor sigma-1 e simulações de bicamadas lipídicas. 2016. 109 p. Tese (Doutorado em Físico-Química) Instituto de Química de São Carlos. Universidade de São Paulo. São Carlos. 2016.
54. LIPINSKI, C.F., Estudo de Relações Quantitativas Estrutura-Atividade de chalconas análogas à combretastatina A4. 2014. 80 p. Dissertação (Mestrado em Físico-Química). Instituto de Química de São Carlos. Universidade de São Paulo. São Carlos. 2014.
55. CHIARI, L.P.A., Estudos de Relações Quantitativas Estrutura-Atividade de Antagonistas do Receptor Sigma-1. 2017. 84 p. Tese (Doutorado em Físico-Química) Instituto de Química de São Carlos. Universidade de São Paulo. São Carlos. 2016.
56. SHRIVER, D. F., ATKINS, P. W., Química Inorgânica, 3ª Ed. **Bookman**: Porto Alegre, v.2, 2003, 1052p.

- 57.KETTLE, S.F.A., Physica Inorganic Chemistry A Coordination Chemistry Approach. **Spektrum Academic Publishers.**, Londres., v. 2, 1996. 506p.
- 58.TODESCHINI, R.; CONSONNI, V. Descriptors from Molecular Geometry. **Handbook of Chemoinformatics.** V. 2, p.414-425 , 2008.
- 59.FUNAR-TIMOFEI, S.; BOROTA, A.; CRISAN, L. Combined molecular docking and QSAR study of fused heterocyclic herbicide inhibitors of protein in photosystem II of plants. **Mol Divers.** V. 21,p. 437-454., 2017.
- 60.AHMED,L.; RASULEV, B.; TURABEKOVA, M.; LESZCZYNSKA, D.; LESZCZYNSKI, J. Receptor- and ligand-based study of fullerene analogues: comprehensive computational approach including quantum-chemical, QSAR and molecular docking simulations. **Organic Biomolecular Chemistry.** v. 35, p.5798-808. 2013.
- 61.POURBASHEER, E.; RIAHI, S.; GANJALI, M.R.; NOROUZI, P. QSAR study on melancortin-4 receptors by support vector machine. **European Journal of Medicinal Chemistry.** v. 45, p.1087-1093, 2010.
- 62.PUZYŃ, T.; LESZCZYŃSKI, M.T. Recent Advances in QSAR Studies: Methods and Applications. **Springer Netherlands**, Frankfurt., v. 1., p.427., 2010.
- 63.DEVINYAK, O., HAVRYLYUK, D., LESYK, R., 3D-MoRSE Descriptors Explained. **Journal of Molecular Graphics and Modeling.**, v. 54, p. 194-203, 2014.
- 64.GALVEZ, J.; GARCIA, R.; SALABERT, M.T.; SOLER, R. Charge Indexes. New Topological Descriptors. **J. Chem. Inf. Comput.** v. 34, p. 520-525, 1994.

ANEXO I

Compostos derivados de Arilpiperazinas com atividade biológica comprovada frente ao receptor 5-HT_{2a} e os respectivos valores de pIC₅₀.

Composto	Estrutura	pIC ₅₀
1		7,51
2		7,57
3		7,41
4		7,73
5		7,32
6		7,68

7		7,84
8		7,59
9		7,26
10		7,99
11		6,86
12		7,70

13		8,29
14		7,89
15		7,77
16		7,68
17		7,71
18		7,52

19		7,15
20		7,24
21		8,06
22		6,97
23		7,04
24		6,81
25		7,12

26		6,98
27		6,88
28		7,34
29		6,66
30		7,09
31		7,53
32		7,86

33		7,88
34		7,91
35		7,89
36		6,87
37		6,77
38		6,59

39		7,63
40		7,20
41		7,82
42		7,15
43		7,68
44		7,62
45		7,72

46		7,40
47		7,19
48		7,42
49		6,93
50		7,27
51		7,51
52		6,85

53		7,91
54		7,56
55		7,54
56		7,10
57		7,04
58		7,01

59		7,17
60		7,70
61		7,49
62		7,49
63		7,55
64		7,39
65		7,32

66		6,62
67		7,72
68		6,61
69		6,69
70		7,62
71		7,24
72		7,23

73		6,07
74		5,99
75		7,01
76		7,28
77		6,43
78		7,25
79		6,75

80		7,20
81		6,23
82		6,90
83		6,77
84		6,73
85		6,69
86		6,98

87		6,71
88		6,24
89		6,60
90		7,11
91		6,71
92		6,49
93		7,15

94		7,03
95		7,57
96		6,50
97		6,74
98		6,56
99		6,75
100		6,97
101		6,83

102		8,01
103		6,90
104		6,64
105		6,63
106		6,53