

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE QUÍMICA DE SÃO CARLOS

ALDINEIA PEREIRA DA SILVA

**Técnicas computacionais aplicadas a uma série de antagonistas do receptor  
5-HT<sub>6</sub> - potencial alvo para a doença de Alzheimer**

São Carlos  
2021

ALDINEIA PEREIRA DA SILVA

**Técnicas computacionais aplicadas a uma série de antagonistas do receptor  
5-HT<sub>6</sub> - potencial alvo para a doença de Alzheimer**

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

Área de concentração: Físico-Química

Orientador: Prof. Dr. Albérico Borges Ferreira da Silva

São Carlos  
2021

## Exemplar revisado

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

*Ficha Catalográfica elaborada pela Seção de Referência e Atendimento ao Usuário do SBI/IQSC*

Silva, Aldineia Pereira da

Técnicas computacionais aplicadas a uma série de antagonistas do receptor  
5-HT<sub>6</sub> - potencial alvo para a doença de Alzheimer / Aldineia Pereira da Silva.

— São Carlos, 2021.

140 f.

Tese (Doutorado em Físico-Química) — Instituto de Química de São Carlos /  
Universidade de São Paulo, 2021.

Orientador: Prof. Dr. Albérico Borges Ferreira da Silva

1. Doença de Alzheimer . 2. Receptor 5-HT<sub>6</sub>. 3. QSAR. 4. Planejamento de  
fármacos. 5. Quimiometria. I. Título.

*Sonia Alves - CRB: 4280/8*



## A corrida da vida

(Bráulio Bessa)

Na corrida dessa vida  
é preciso entender  
que você vai rastejar,  
que vai cair, vai sofrer  
e a vida vai lhe ensinar  
que se aprende a caminhar  
e só depois a correr.

A vida é uma corrida  
que não se corre sozinho.  
E vencer não é chegar,  
é aproveitar o caminho  
sentindo o cheiro das flores  
e aprendendo com as dores  
causadas por cada espinho.

Aprenda com cada dor,  
com cada decepção,  
com cada vez que alguém  
lhe partir o coração.  
O futuro é obscuro  
e às vezes é no escuro  
que se enxerga a direção.

Aprenda quando chorar  
e quando sentir saudade,  
aprenda até quando alguém  
lhe faltar com a verdade.  
Aprender é um grande dom.  
Aprenda que até o bom  
vai aprender com a maldade.

Aprender a desviar  
das pedras da ingratidão,  
dos buracos da inveja,  
das curvas da solidão,  
expandindo o pensamento  
fazendo do sofrimento  
a sua maior lição.

Sem parar de aprender,  
aproveite cada flor,  
cada cheiro no cangote,  
cada gesto de amor,  
cada música dançada  
e também cada risada,  
silenciando o rancor.

Experimente o mundo,  
prove de todo sabor,  
sinta o mar, o céu e a terra,  
sinta o frio e o calor,  
sinta sua caminhada  
e dê sempre uma parada  
pelo caminho que for.

Pare, não tenha pressa,  
não carece acelerar,  
a vida já é tão curta,  
é preciso aproveitar  
essa estranha corrida  
que a chegada é a partida  
e ninguém pode evitar!

Por isso é que o caminho  
tem que ser aproveitado,  
deixando pela estrada  
algo bom pra ser lembrado,  
vivendo uma vida plena,  
fazendo valer a pena  
cada passo que foi dado.

Aí sim, lá na chegada,  
onde o fim é evidente,  
é que a gente percebe  
que foi tudo de repente,  
e aprende na despedida  
que o sentido da vida  
é sempre seguir em frente.

## DEDICATÓRIA

Dedico esta tese aos meus pais, Maria Eunice e José Carlos, e à minha madrinha, Sebastiana Pereira (a quem eu tenho como segunda mãe) por pavimentarem a estrada pela qual eu passei, suavizando minha caminhada a cada passo dado. A vocês, deixo aqui, o meu muito obrigado.

## AGRADECIMENTOS

Ao longo da minha jornada, muitas pessoas contribuíram para suavizar o meu caminhar. Tenho convicção que o apoio, ensinamento e incentivo a mim dedicado contribuíram para que eu suportasse os fardos dessa luta, por isso, deixo aqui os meus sinceros agradecimentos:

- Aos meus pais, José Carlos e Maria Eunice, pelo maior e melhor presente que um pai e uma mãe poderiam oferecer a um filho - EDUCAÇÃO. Sei do esforço diário de vocês para que eu galgasse os maiores degraus da escada da vida e sei que a luta de vocês foi tão ou maior do que a minha. Por isso, o meu muito obrigado é insuficiente para expressar o quanto eu sou agradecida;
- À minha madrinha, Sebastiana Pereira, por todo apoio, carinho, por todos os ensinamentos para que eu me tornasse uma pessoa melhor, pelas orações, pelas vibrações com minhas conquistas e por me ensinar a nunca baixar a cabeça diante das dificuldades. Muito obrigada por fazer parte da minha vida, a senhora é luz, te amo;
- A todos os meus irmãos, Fabiano, Adriano, Aldemir, meu Dodinho, Vânia, a quem eu amo independente de qualquer coisa, e Júnior, o caçula, pela torcida;
- Um agradecimento especial à minha irmã querida, amiga e incentivadora incondicional, Aldineide Nunes. A nossa relação é muito maior que o DNA que nos une, é de amor profundo, é de leveza, é de amizade, e eu sei que as minhas conquistas são como se fossem as suas e vice-versa. Muito obrigada por sua presença constante em minha vida, pelo seu apoio, pela sua alegria diante das minhas conquistas e por tudo que você representa, te amo muito.
- Um agradecimento especial a toda família Pereira, por todos os momentos de alegria que compartilhamos;
- Aos meus primos, Netinho e Paulinha, pelo incentivo, carinho e por tudo que vocês fizeram por mim. Sou sortuda por tê-los em minha vida, muito obrigada.
- Não poderia deixar de agradecer a duas primas muito especiais em minha vida, Márcia e Marcela. Vocês não sabem, mas em dias mais difíceis, vocês me fizeram rir, me passaram

confiança, me apoiaram e isso tudo só engrateceu a nossa relação. Meu coração é de gratidão e amor por vocês.

- Quando o tédio da pós-graduação estava em níveis insuportáveis, o meu refúgio foi na casa da minha tia Valdelice. Obrigada por todas as vezes que me recebeu em sua casa e preparou todas as comidas que eu gosto de comer. Esse carinho é impagável;
- Agradeço também à minha tia Josy, exemplo de força, garra e determinação em busca de seus objetivos, pessoa a quem eu me espelho;
- À Joice, querida tia, que sempre me tratou com carinho e sempre me recebeu com muito amor em sua casa. Sou grata por todas às vezes que atendeu os meus pedidos para cozinhar aquelas comidas que só nós, nordestinas, gostamos.
- Ao meu amigo, Anderson Arandas, por todo carinho, por todo cuidado, por sempre está comigo, por sempre fazer questão de me mostrar os meus pontos positivos e levantar minha cabeça em dias mais complicados. Você é uma benção em minha vida, muito obrigada por todos esses anos de amizade.
- Ao meu amigo, Eri, por tudo que passamos juntos. Obrigada, meu querido, pela sua amizade.
- À família São Carlos, Gisele, Aline, Lílian e Amandita por todos os momentos que compartilhamos. Sortuda sou eu por tê-las como amigas.
- Um Agradecimento especial à Amanda (Amandita) não só pela amizade, mas também pelas contribuições, correções e sugestões para melhorar a minha tese de doutorado.
- À minha amiga Virgínia, por todos os momentos de incentivo, por contribuir para o meu crescimento pessoal e também profissional, pela amizade, carinho e preocupação comigo. Obrigada minha amiga, te amo muito e te tenho como exemplo de bondade e dedicação aos seus amigos.
- Aos colegas do laboratório, Gênisson, Nati, Anny, Ana, Júlia e Herison pelos momentos de descontração que compartilhamos.

- À Laise, pela parceria no desenvolvimento de vários trabalhos nos últimos tempos e por todos os momentos de incentivos que nos demos;
- Ao professor Guedmiller Souza da Universidade Federal de Uberlândia pela contribuição para o desenvolvimento de parte dessa pesquisa;
- Ao meu orientador, Albérico, por todos os “puxões de orelhas”, palavras de incentivos e às vezes palavras um pouco mais duras, mas que contribuíram para o meu crescimento, sem deixar de lado os momentos de descontração. Obrigada, professor;
- Um agradecimento muito mais que especial a todo o povo Brasileiro, que por meio dos impostos pagos custearam minha estadia ao longo dos 54 meses de bolsas pagas por meio da CAPES. Sei que a grande maioria de vocês nunca teve a oportunidade de entrar numa universidade. Talvez, os filhos e quem sabe até mesmo os netos de vocês não terão o mesmo grau instrucional que eu. Mesmo enfrentando diversas dificuldades para chegar até esse ponto, tenho consciência que sou privilegiada nessa sociedade tão desigual.
- O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES)- Código de Financiamento 001.

O correr da vida embrulha tudo.  
A vida é assim: esquenta e esfria,  
Aperta e daí afrouxa,  
Sossega e daí desinquieta.  
O que ela quer da gente é coragem.  
**(Guimarães Rosa)**

## RESUMO

Da Silva, A. P. **Técnicas computacionais aplicadas a uma série de antagonistas do receptor 5-HT<sub>6</sub> - potencial alvo para a doença de Alzheimer**. 2021. Tese (Doutorado) – Instituto de química de São Carlos, Universidade de São Paulo, São Carlos, 2021.

Os receptores acoplados à proteína G (GPCR) são alvos de aproximadamente um terço dos medicamentos atualmente comercializados no mundo. Nos últimos anos, o receptor 5-HT<sub>6</sub> (família GPCR) vem atraindo atenção devido à sua localização em áreas do cérebro envolvendo memória e aprendizagem, o colocando como um dos principais alvos na busca por novas abordagens terapêuticas anti-Alzheimer, uma vez que os medicamentos em uso não são eficazes. Com o número crescente de idosos acometidos por essa enfermidade, especialmente no Brasil, assomado à inefetividade dos tratamentos disponíveis, é urgente que se investiguem novas terapias medicamentosas para aplacar esse problema. Nesta linha, acredita-se que a administração de antagonistas do receptor 5-HT<sub>6</sub> pode contribuir para mitigar os problemas relacionados à perda de memória e à deterioração da aprendizagem, efeitos esses associados à doença de Alzheimer. Para aumentar as chances de se obter terapias medicamentosas eficazes diante das perdas amnésicas decorrentes dessa enfermidade, planejou-se novos antagonistas do receptor 5-HT<sub>6</sub> lançando-se mão de abordagem *in Silico*. Os novos compostos foram submetidos aos modelos QSARs aqui gerados e promissores valores de atividade biológica foram obtidos. Esses achados, somados ao ótimo perfil ADME/Tox alcançado via aplicação de ferramentas computacionais, qualificam esses potenciais antagonistas para futuras abordagens experimentais, o que poderiam confirmar suas características drogáveis.

**Palavras-Chave:** Doença de Alzheimer, receptor 5-HT<sub>6</sub>, QSAR, planejamento de fármacos.

## ABSTRACT

Da Silva, A. P. **Computational techniques applied to a series of 5-HT<sub>6</sub> receptor antagonists - target potential for Alzheimer's disease**. 2021. Thesis (Doctoral)- Instituto de química de São Carlos, Universidade de São Paulo, São Carlos, 2021.

G Protein-coupled receptors (GPCRs) are targets for approximately one-third of the drugs currently marketed worldwide. Over recent years, the 5-HT<sub>6</sub> receptor (GPCR family) has been attracted attention due to locations in brain regions involving learning and memory processes, reason why it became one of the main targets in the race for the discovery of new anti-Alzheimer therapeutic approaches once the ongoing drugs are not effective. With the growing number of elderly people affected by this disease, especially in Brazil, in addition to the ineffectiveness of available treatments, it is urgent that new pharmacological treatments be investigated to mitigate this problem. Addressing this problem, it is believed that the administration of 5-HT<sub>6</sub> receptor antagonists might mitigate problems related to memory loss and impaired learning, effects associated with Alzheimer's disease. To increase the chances of obtaining effective medical countermeasures against amnesic losses due to this disease, new antagonists of the 5-HT<sub>6</sub> receptor were designed using *in Silico* approach. The new compounds were submitted to the QSARs models and excellent values of biological activity were obtained. Finally, the ADME/Tox profile of these compounds was ascertained by the applications of computational tools and the outcomes achieved qualify them for subsequently experimental procedures, which could confirm the antagonistic effects of these compounds as well as their drug-like properties.

**Keywords:** Alzheimer's disease, 5-HT<sub>6</sub> receptor, QSAR, Drug design.

## LISTA DE FIGURAS

<b>Figura 1-</b> Países com planos de ação em atenção à saúde do idoso .....	22
<b>Figura 2-</b> Os oligômeros A $\beta$ se agregam formando placas $\beta$ -amiloides. Estas, por sua vez, entram na membrana e abrem os canais de Ca <sup>2+</sup> e de outros íons, causando excitabilidade nas células neuronais e, posteriormente, apoptose neuronal.....	24
<b>Figura 3</b> - Estrutura química da desferroxiamina .....	25
<b>Figura 4-</b> Fármacos usados no tratamento da DA. <b>a)</b> galantamina, <b>b)</b> rivastigmina, <b>c)</b> tacrina, <b>d)</b> donepezil .....	26
<b>Figura 5-</b> Mecanismo da hipótese colinérgica- a colina acetiltransferase sintetiza acetilcolina que, por sua vez, se liga a dois receptores (nicotínicos e muscarínicos). Por meio da ação da enzima acetilcolinesterase, a acetilcolina é degradada, produzindo colina.....	27
<b>Figura 6</b> - Mecanismo da hipótese glutamatérgica- o glutamato age por meio de três tipos de receptores, NMDA, AMPA e Cainato. Uma vez que os receptores NMDA são ativados de maneira exagerada, íons Ca <sup>2+</sup> são liberados demasiadamente. Os receptores AMPA e Cainato também estimulam a entrada de íons Ca <sup>2+</sup> , ao mesmo tempo que afastam os íons Mg <sup>2+</sup> . Este por sua vez, atua bloqueando o entorno da membrana dentro do canal NMDA. Como resultado, o glutamato se liga a outra subunidade do receptor NMDA, atuando junto com a glicina que está ligada a outra subunidade. Toda essa cascata de evento resulta em mais liberação de íons Ca <sup>2+</sup> , provocando a morte do neurônio .....	28
<b>Figura 7-</b> Estrutura química da memantina .....	29
<b>Figura 8</b> - Representação dos experimentos que culminaram na descoberta da barreira hematoencefálica. <b>a)</b> Injeção do corante na corrente sanguínea; <b>b)</b> Injeção do corante no SNC .....	30
<b>Figura 9</b> - Diagrama esquemático da BHE - as junções oclusivas, células do capilar sanguíneo e membrana basal, formam uma barreira protetora que impede a entrada de corpos estranhos no sistema nervoso central.....	31
<b>Figura 10-</b> Estrutura geral de um receptor 7TM.....	34
<b>Figura 11-</b> Mecanismo geral de interação dos GPCRs com as subunidades da proteína G - na etapa inicial, o receptor interage com o ligante e ativa a proteína G. A subunidade $\alpha$ se desliga da GDP para se ligar à GTP, dissociando do dímero $\beta\gamma$ . Ao finalizar a transdução a subunidade $\alpha$ retorna ao seu estado basal. ....	35
<b>Figura 12-</b> Estrutura química do Ro 04-6790 - antagonista do receptor 5-HT <sub>6</sub> . ....	36

<b>Figura 13</b> - Características dos elementos estruturais dos antagonistas do 5-HT <sub>6</sub> R. Vermelho: <b>PI</b> , Verde: <b>AR</b> ; Azul: <b>HBA</b> ; Magenta: <b>HYD</b> .....	37
<b>Figura 14</b> - Processo de ancoramento molecular- <b>a)</b> sítio ativo; <b>b)</b> ligante; <b>c)</b> encaixe do ligante no sítio ativo .....	39
<b>Figura 15</b> - Representação do funcionamento de um AG - atribui-se aleatoriamente valores a cada ângulo torcional (gene), sendo o conjunto desses valores representado por um cromossomo.....	41
<b>Figura 16</b> - As partículas em um sistema molecular sofrem influência das forças de outras partículas. Em DM, todas as forças, termos ligados e termos não-ligados, são acopladas em uma função matemática denominada de campo de força. ....	43
<b>Figura 17</b> - Etapas percorridas em uma simulação usando DM.....	45
<b>Figura 18</b> - Condições periódicas de contorno - nesse tipo de simulação, apenas a caixa central é simulada. Embora o sistema seja delimitado, as partículas não interagem com as paredes da caixa, e assim consegue-se simular o ambiente fisiológico de uma proteína, por exemplo.....	47
<b>Figura 19</b> - Etapas seguidas para a construção de modelos QSAR- seleção de um conjunto de compostos, cálculo e seleção de descritores, construção e validação do modelo QSAR, planejamento de compostos e predição da atividade biológica.....	51
<b>Figura 20</b> - Modificações estruturais realizadas em um composto hipotético .....	52
<b>Figura 21</b> - Estrutura geral de uma Rede Neural adaptada para construção de modelos QSAR .....	60
<b>Figura 22</b> - Simulação de <i>Dinâmica</i> molecular- <b>a)</b> Representação de todo o sistema: membrana, moléculas de água, receptor e ligante; <b>b)</b> representação do sistema sem a membrana POPC; <b>c)</b> visualização do receptor, ligante e membrana; <b>d)</b> representação do sistema sem as moléculas de água.....	69
<b>Figura 23</b> - Alinhamento realizado entre a sequência primária do receptor 5-HT <sub>6</sub> e alguns outros alvos GPCRs. Nesse alinhamento, a região em vermelho são os resíduos conservados, aqueles que estão na mesma posição entre todos os alvos biológicos.....	72
<b>Figura 24</b> - Regiões conservadas e variáveis do receptor 5-HT <sub>6</sub> .....	73
<b>Figura 25</b> - Modelo construído e suas regiões conservadas .....	73
<b>Figura 26</b> - Qualidade estereoquímica do modelo tridimensional do receptor 5-HT <sub>6</sub> .....	74
<b>Figura 27</b> – Comparação do modelo refinado por meio do cálculo da energia DOPE .....	75
<b>Figura 28</b> - Avaliação externo do modelo do receptor 5-HT <sub>6</sub> depois de refinado via método DM.....	76

<b>Figura 29</b> - Contribuição de cada descritor para o vetor de regressão relacionado ao modelo PLS .....	79
<b>Figura 30</b> - Detecção de outlier no modelo QSAR-2D .....	80
<b>Figura 31</b> - Análise estatística da robustez interna do modelo QSAR-2D obtido.....	80
<b>Figura 32</b> - teste de y-randomização aplicado ao modelo .....	81
<b>Figura 33</b> - Predição dos valores de $pK_i$ para os conjuntos treinamento e teste.....	82
<b>Figura 34</b> - Estrutura molecular dos três compostos mais ativos do conjunto de dados .....	83
<b>Figura 35</b> - Diferenças estruturais entre o composto mais ativo e menos ativo do conjunto de dados.....	84
<b>Figura 36</b> - Orbitais LUMOs dos compostos mais ativos .....	85
<b>Figura 37</b> - Valores de $pK_i$ experimentais <i>versus</i> preditos obtidos a partir do modelo CoMFA .....	87
<b>Figura 38</b> - Análises das contribuições eletrostáticas e estéricas para o composto 25. <b>a)</b> contribuições eletrostáticas positivas (azul) e negativas (vermelho) para a atividade biológica, respectivamente; <b>b)</b> contribuições estereoquímicas favoráveis (verde) e desfavoráveis (amarelo) para a atividade biológica, respectivamente .....	88
<b>Figura 39</b> - Análises das contribuições eletrostáticas e estéricas para o composto 47.....	89
<b>Figura 40</b> - Análises das contribuições eletrostáticas e estéricas para o composto 15.....	89
<b>Figura 41</b> - Valores de $R^2$ obtidos para o conjunto total e treinamento .....	91
<b>Figura 42</b> - Valores de $R^2$ obtidos para o conjunto teste e validação .....	92
<b>Figura 43</b> - Erro quadrado médio obtido para os conjuntos treinamento, teste e validação em função do número de iterações .....	93
<b>Figura 44</b> - Detecção de outlier para todos os compostos do conjunto 2.....	94
<b>Figura 45</b> - Verificação da consistência interna do modelo ANN-QSAR .....	95
<b>Figura 46</b> - Valores de $pK_i$ experimental <i>versus</i> preditos pelo modelo ANN-QSAR para os compostos selecionados do ChEMBL.....	95
<b>Figura 47</b> - Dimensionalidade das representações estruturais - <b>a)</b> descritores 2D e <b>b)</b> descritores 3D .....	97
<b>Figura 48</b> - exemplo de uma matriz de distância para a molécula isopentano .....	97
<b>Figura 49</b> - Compostos que apresentaram valor do descritor Jhetz igual a 1 .....	98
<b>Figura 50</b> - Compostos com valor do descritor BEHm2 igual a 6, mas com baixo valor de $pK_i$ .....	99

<b>Figura 51</b> - Análises dos descritores GATS3p e GATS1p para os compostos com maiores valores de $pK_i$ .....	100
<b>Figura 52</b> - compostos que apresentaram valores de RDF025u na faixa de 35 a 38.....	101
<b>Figura 53</b> - Compostos que apresentaram valores de RDF025u acima de 49.....	101
<b>Figura 54</b> - representação dos orbitais HOMO para os compostos mais ativos ( <b>62</b> e <b>70</b> ), os com efeito biológico moderado ( <b>63</b> e <b>48</b> ) e os menos ativos ( <b>67</b> e <b>81</b> ) .....	103
<b>Figura 55</b> - Tempo de simulação <i>versus</i> RMSD para o receptor 5-HT <sub>6</sub> complexado ao ligante C15 embebida em seu ambiente fisiológico .....	104
<b>Figura 56</b> - Principais tipos de interações entre o <b>C15</b> e o sítio ativo do receptor 5-HT <sub>6</sub> : representação 3D (imagem à esquerda) e diagrama bidimensional (imagem à direita).....	105
<b>Figura 57</b> - Possível mecanismo molecular do bloqueio do receptor 5-HT <sub>6</sub> .....	105
<b>Figura 58</b> - Comparação entre os modos de ligação dos compostos <b>C15</b> (estrutura em magenta) e <b>C25</b> (estrutura em azul) - <b>a</b> ) representação bidimensional dos principais resíduos de aminoácidos envolvidos na interação com o <b>C25</b> ; <b>b</b> ) ligação de hidrogênio entre o resíduo LEU <sup>183</sup> e o <b>C25</b> ; <b>c</b> ) comparação entre as poses de ligação entre os compostos <b>C15</b> e <b>C25</b> ; <b>d</b> ) Diferença no modo de interação entre os compostos e o resíduo GLN <sup>292</sup> . <b>Tracejado em magenta</b> : ligação de hidrogênio; <b>tracejado em verde</b> : ligação de hidrogênio do tipo CH...π. ....	107
<b>Figura 59</b> - Principais interações entre o composto <b>70</b> e o sítio ativo do receptor 5-HT <sub>6</sub> – <b>a</b> ) diagrama bidimensional, <b>b</b> ) representação tridimensional .....	108
<b>Figura 60</b> - Principais tipos de interações entre os resíduos de aminoácidos e o composto <b>C67</b> - <b>a</b> ) representação bidimensional, <b>b</b> ) representação tridimensional.....	110
<b>Figura 61</b> - Simulação de DM para o complexo receptor-ligante ( <b>C4704</b> ) embebido em seu ambiente fisiológico. No canto esquerdo do gráfico tem-se uma comparação entre os RMSDs para as simulações com <b>C15</b> (gráfico em vermelho e composto <b>C4704</b> (representação em azul) .....	113
<b>Figura 62</b> - Simulação de <i>Docking</i> molecular entre o receptor 5-HT <sub>6</sub> e composto <b>C4704</b> . <b>a</b> ) representação bidimensional dos resíduos de aminoácidos que interagem com o <b>C4704</b> ; <b>b</b> ) principal tipo de interação entre o composto <b>C4704</b> e o resíduo ARG <sup>174</sup> .....	114

## LISTA DE TABELAS

<b>Tabela 1</b> - Parâmetros estatísticos analisados para validar internamente o modelo QSAR-2D .....	58
<b>Tabela 2</b> - Conjunto 1: estrutura e atividade biológica dos antagonistas do receptor 5-HT <sub>6</sub> ..	61
<b>Tabela 3</b> - Conjunto 2: estrutura e valores de atividade biológica dos antagonistas do receptor 5-HT <sub>6</sub> selecionados para a construção do modelo ANN (nesse conjunto, os compostos do conjunto teste estão sublinhados e os do conjunto validação estão destacados em azul) .....	63
<b>Tabela 4</b> - Estrutura molecular dos compostos usados para verificar o poder preditivo do modelo ANN-QSAR .....	66
<b>Tabela 5</b> - Parâmetros estatísticos analisados para escolher os descritores que participarão do modelo QSAR-2D .....	77
<b>Tabela 6</b> - Definição dos descritores selecionados e classe para a qual eles pertencem .....	78
<b>Tabela 7</b> - Valores dos parâmetros estatísticos analisados para a seleção do melhor modelo	78
<b>Tabela 8</b> - Resultados estatísticos para os modelos “sem focagem” e melhor focagem encontrada.....	86
<b>Tabela 9</b> - Resíduos calculados para os compostos do conjunto teste.....	87
<b>Tabela 10</b> - Parâmetros estatísticos analisados para a seleção dos descritores.....	90
<b>Tabela 11</b> - Definição dos descritores selecionados para a construção do modelo ANN -QSAR .....	90
<b>Tabela 12</b> - Valores de pK <sub>i</sub> preditos para todos os compostos do conjunto teste.....	94
<b>Tabela 13</b> - Valores de pK <sub>i</sub> preditos pelo modelo ANN-QSAR.....	96
<b>Tabela 14</b> -compostos planejados a partir das modificações estruturais nas moléculas C15, C47 e C25 .....	111
<b>Tabela 15</b> - Compostos propostos a partir das análises do modelo ANN-QSAR .....	115
<b>Tabela 16</b> - Estimativa do perfil ADEME/TOX dos novos compostos .....	117
<b>Tabela 17</b> - Predição da acessibilidade sintética para os novos compostos .....	119

## LISTA DE ABREVIATURAS E SIGLAS

- ACh** - Acetilcolina
- ACHe** - Acetilcolinesterase
- AD** - Doença de Alzheimer
- ADI** - Alzheimer's Disease International
- ADME/Tox** - Absorção, Distribuição, Metabolismo e Excreção/Toxicidade
- AG** - algoritmo genético
- AMPA** -  $\alpha$ -amino-3-hidróxi-5-metil-4-isoxazolepropiónico
- ANN** - Artificial Neural Network
- AO** - Oligonucleotídeo Anti-sentido
- APP** - Amyloid Precursor Protein
- BHE** - Barreira hematoencefálica
- CAT** - Acetiltransferase
- CoMFA** - Comparative Molecular Field. Analysis
- DFO** - Desferroxiamina
- DFT** - Density Functional Theory
- DM** - Dinâmica Molecular
- DOPE** - Discrete Optimized Protein Energy
- ECL** - Extracelular
- FDA** - Food and Drug Administration
- FHP** - Filamentos helicoidais pareados
- GDP** - Guanosina difosfato
- GPCR** - Protein Coupled Receptor
- GTP** - Guanosina trifosfato
- HOMO** - Highest Occupied Molecular Orbital
- ICL** - Intracelular
- JA** - Junções apertadas
- JEA** - Junções endoteliais aderentes
- LBDD** - Ligand-Based Drug Design
- LUMO** - Lowest Unoccupied Molecular Orbital
- LM** - Levenberg-Marquardt
- MLP** - Multilayer Perceptron

**MSE** - Mean-Squared Error

**NCBI**- National Center for Biotechnology Information

**NFT** - Emaranhados neurofibrilares

**NMDA** - N-Metil-D-Aspartato

**OECD** - Organisation de Coopération et de Développement Economiques Organisation for Economic Co-operation and Development

**PDB** - Protein Data Bank

**PLS** - Parcial Least Squares

**POPC**- 1-plalmiloil-2-oleoil-sn-glicerol-3-fosfocolina

**PPW** - Protein Preparation Wizard

**QSAR** - Quantitative Structure-Activity Relationships

**RMN** - Ressonância Magnética Nuclear

**SAR** -Structure-Activity Relationships

**SBDD** - Structure-Based drug design

**SNC** - Sistema nervoso central

**SVM** - vetores de suporte de máquina

**THS** - High-Throughput Screening

**TIP3P** - Transferable Intermolecular Potentials

**5-HT** - Receptores da Serotonina

**7TM** -Transmembranar

**βA** - Beta-amilóide

# Sumário

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>21</b>
1.1	Hipótese da cascata amiloide e o papel da proteína tau na doença de Alzheimer .....	23
1.2	Hipótese Colinérgica.....	26
1.3	Hipótese glutamatérgica .....	27
1.4	A barreira hematoencefálica- um obstáculo a ser contornado .....	29
1.5	Receptor 5-HT <sub>6</sub> .....	34
<b>2</b>	<b>APLICAÇÃO DE MÉTODOS COMPUTACIONAIS NO PLANEJAMENTO DE FÁRMACOS .....</b>	<b>38</b>
2.1	Abordagem molecular focada no receptor .....	38
2.1.1	Docking Molecular .....	42
2.1.2	Dinâmica Molecular .....	49
2.1.3	Predição das coordenadas tridimensionais de uma proteína .....	51
2.2	Abordagem molecular focada no ligante .....	51
2.3	Relação Quantitativa Estrutura-Atividade-QSAR .....	51
<b>3</b>	<b>OBJETIVOS.....</b>	<b>54</b>
3.1	Objetivos Específicos .....	54
<b>4</b>	<b>METODOLOGIA.....</b>	<b>55</b>
4.1	Construção das coordenadas tridimensionais do receptor 5-HT <sub>6</sub> .....	55
4.2	Construção dos modelos QSARs (2D e 3D).....	56
4.2.1	PLS (Partial Least Squares).....	56
4.2.2	Análises Comparativa de Campos Moleculares (CoMFA).....	58
4.2.3	Redes Neurais Artificiais.....	59
4.3	Conjunto de dados .....	60
4.4	Simulação de Docking molecular .....	68
4.5	Simulações de Dinâmica molecular.....	69
4.6	Procedimentos.....	70
<b>5</b>	<b>RESULTADOS E DISCUSSÕES.....</b>	<b>72</b>
5.1	Construção das coordenadas tridimensionais do receptor 5-HT <sub>6</sub> .....	72
5.2	Geração dos modelos QSAR .....	77

5.2.1	Primeiro conjunto de compostos .....	77
5.2.1.1	Análise dos descritores selecionados para a construção do modelo QSAR-2D.....	82
5.2.1.2	Análises QSAR-3D .....	86
5.2.2	Segundo conjunto de compostos .....	90
5.2.3	Análise dos descritores selecionados.....	96
5.3	Análises da simulação de Dinâmica molecular .....	103
5.4	Análises das simulações de Docking molecular .....	104
5.5	Planejamento de novos compostos .....	111
5.6	Avaliação do perfil ADME/TOX dos novos compostos .....	117
<b>6</b>	<b>CONCLUSÃO.....</b>	<b>120</b>

## 1 INTRODUÇÃO

A doença de Alzheimer (DA) é uma patologia neurodegenerativa, cuja característica principal está associada à deterioração crônica das funções cognitivas.<sup>1,2</sup> Em um quadro de doenças relacionadas à demência, a DA não só é a enfermidade mais frequente, como também a principal causa de morte entre os idosos.<sup>3</sup>

Os prejuízos decorrentes da demência ocorrem devido aos danos ou destruição das células neuronais presentes no cérebro.<sup>4</sup> Esses prejuízos são relativos à perda progressiva da memória, disruptura na fala, desorientação espaço-tempo, raciocínio, concentração, aprendizado, movimento, e tantas outras habilidades.

O prenúncio da doença de Alzheimer ocorreu em 1901 quando Auguste Deter, 51 anos, acompanhada do seu marido, chegou ao consultório do Dr. Alois Alzheimer. O resultado da consulta revelou um quadro flutuante entre melancolia e descontentamento. Ela recordava-se de vários acontecimentos, mas não conseguia lembrar de eventos atuais, esquecendo, inclusive, do seu próprio nome. Esse diagnóstico deixou o Dr. Alois bastante intrigado, pois o exame geral mostrou que Auguste Deter estava saudável e o neurológico também seguiu nessa linha.<sup>5-7</sup>

Cinco anos depois (1906), a senhora Deter foi a óbito, o que possibilitou a investigação detalhada do seu cérebro. As análises revelaram acúmulo de placas entre as células neuronais (conhecidas como beta-amilóide -  $\beta$ A) e uma espécie de emaranhado fibrilar intraneuronal. A partir da revelação desses biomarcadores, descobriu-se a doença de Alzheimer (nome dado em homenagem ao Dr. Alois Alzheimer).<sup>5-7</sup>

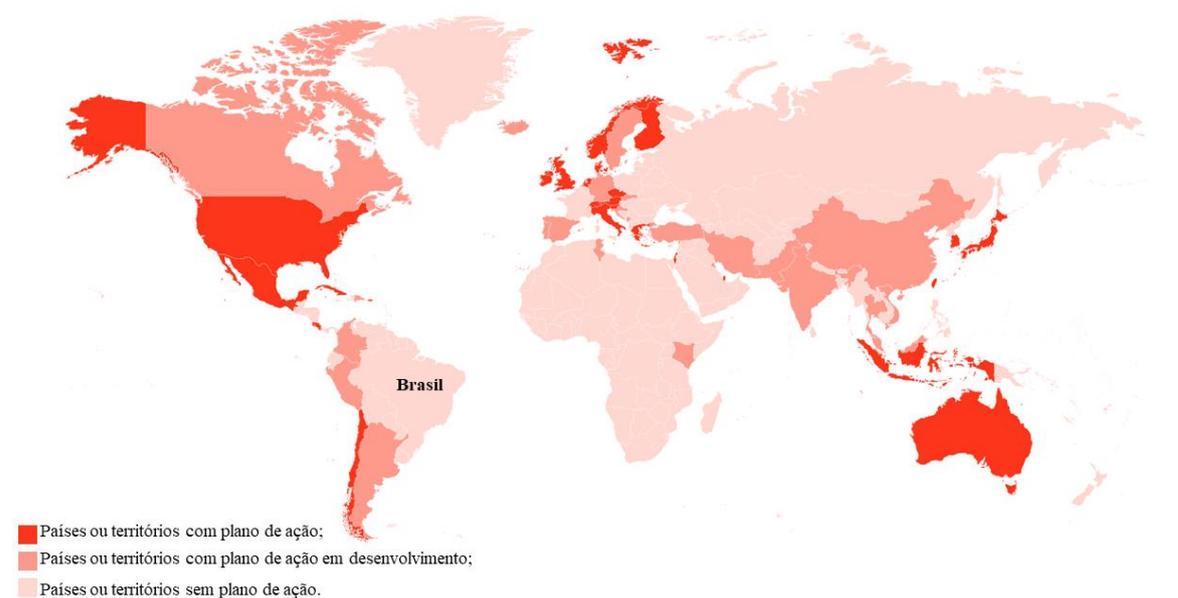
Segundo pesquisas<sup>8-11</sup>, o prelúdio ao aparecimento dos primeiros sintomas visíveis da doença de Alzheimer ocorre 20 anos mais cedo, fase em que surgem as primeiras alterações no cérebro das pessoas acometidas por essa patologia, inicialmente imperceptíveis. Essas modificações se multiplicam com o decorrer do tempo até que os sintomas se tornem visíveis, interferindo na capacidade dos indivíduos de realizarem tarefas diárias simples.

De acordo com a organização Alzheimer's Disease International (ADI)<sup>12</sup> estima-se que mais 50 milhões de pessoas no mundo já foram diagnosticadas com essa neuropatologia, e que a cada 3 segundos, alguém desenvolve algum tipo de demência. Calcula-se que o custo relacionado à demência é superior a US\$ 1 trilhão, e até 2030 este número poderá duplicar. No mundo, entre 2000 e 2016, mortes relacionadas a algum tipo de demência mais que dobraram, saindo da 14ª causa de morte para a 5ª se comparada à doença cardíaca isquêmica, doença pulmonar obstrutiva crônica, hemorragia intracerebral e acidente vascular cerebral isquêmico.<sup>13</sup> Somente

no Brasil<sup>14,15</sup>, estima-se que cerca de 1,6 milhões de pessoas estão acometidas por algum tipo de demência.

Em resposta a esse quadro alarmante, vários países desenvolveram ou estão desenvolvendo planos de ação em atenção à saúde dos idosos (Figura 1).<sup>16</sup> No Brasil, no entanto, esses planos de ações se apresentam de forma tímida ou inexistem, apesar das estatísticas apontarem dados preocupantes – é o 2º país com maior número de casos de demência (1037 por 882-1220 mil habitantes).<sup>13</sup>

**Figura 1-** Países com planos de ação em atenção à saúde do idoso



**Países com planos de ação em atenção à saúde do idoso**

Austrália	Czech Republic	Israel	Macau SAR	Puerto Rico	Switzerland
Austria	Denmark	Italy	Malta	Qatar	TADA Chinese
Chile	Finland	Ireland	Mexico	Republic of Korea	Taipei
Costa Rica	Greece	Japan	Netherlands	Slovenia	UK
Cuba	Indonesia	Luxembourg	Norway	Singapore	USA

**Países com planos de ação em desenvolvimento em atenção à saúde do idoso**

Argentina	China	Dominican Rep.	Kenya	Pakistan	Sweden
Barbados	Colombia	Germany	Kuwait	Peru	Thailand
Bonaire	Croatia	Iceland	Malaysia	Portugal	Tunisia
Brunei	Curacao	India	Mauritius	Spain	Turkey
Canada	Cyprus	Iran	Montenegro	Slovak Rep	Vietnam

Fonte: Alzheimer’s Disease International.<sup>13</sup>

O cérebro de um indivíduo saudável é formado por cerca de 100 bilhões de neurônios, 100 trilhões de sinapses e dezenas de neurotransmissores. Sinais elétricos são gerados nos neurônios, os quais se conectam por meio de sinapses, resultando na liberação de neurotransmissores, estes são os responsáveis por transportar informações químicas para outras células nervosas.<sup>17</sup> A forma como os neurônios se conectam nos levam a entender como os padrões de atividades são formados em nível celular, por exemplo aprender uma nova língua, aprender matemática, química, física, e outras atividades mais simples, como comer, andar, falar, se vestir, dentre outras. No entanto, quando se trata de pessoas acometidas por Alzheimer, o número de sinapses cai abruptamente conforme a doença avança, prejudicando a prática de todas essas atividades.<sup>17</sup>

Para compreender o mecanismo de ação dos medicamentos administrados a pacientes acometidos pelo mal de Alzheimer, é necessário diferenciar dois biomarcadores: neurodegeneração (morte das células neurais que causam a doença) e o declínio cognitivo (manifestação clínica desta patologia).<sup>17</sup> No entanto, os medicamentos em uso não focam na causa, isto é, morte celular, mas sim no melhoramento das funções cognitivas, tais como memória e aprendizagem, conforme será evidenciado nas próximas seções.

### *1.1 Hipótese da cascata amiloide e o papel da proteína tau na doença de Alzheimer*

A hipótese da cascata amiloide, defendida pelo Dr. Alois Alzheimer após estudar o cérebro “*post mortem*” de uma paciente, está associada ao acúmulo de placas  $\beta$ -amiloide em células extraneuronais e à hiperfosforilação da proteína tau, configurada na forma de emaranhados neurofibrilares (NFT).<sup>18,19</sup>

Em neurônios saudáveis, a proteína tau encontra-se nos axônios. No entanto, em indivíduos acometidos por doenças neurodegenerativas, tal como DA, esta proteína é hiperfosforilada, tornando-se insolúvel. Essa hiperfosforilação se estende por todo o corpo celular e se apresenta sob a forma de filamentos helicoidais pareados (FHP), principais componentes dos NFTs.<sup>18,19</sup>

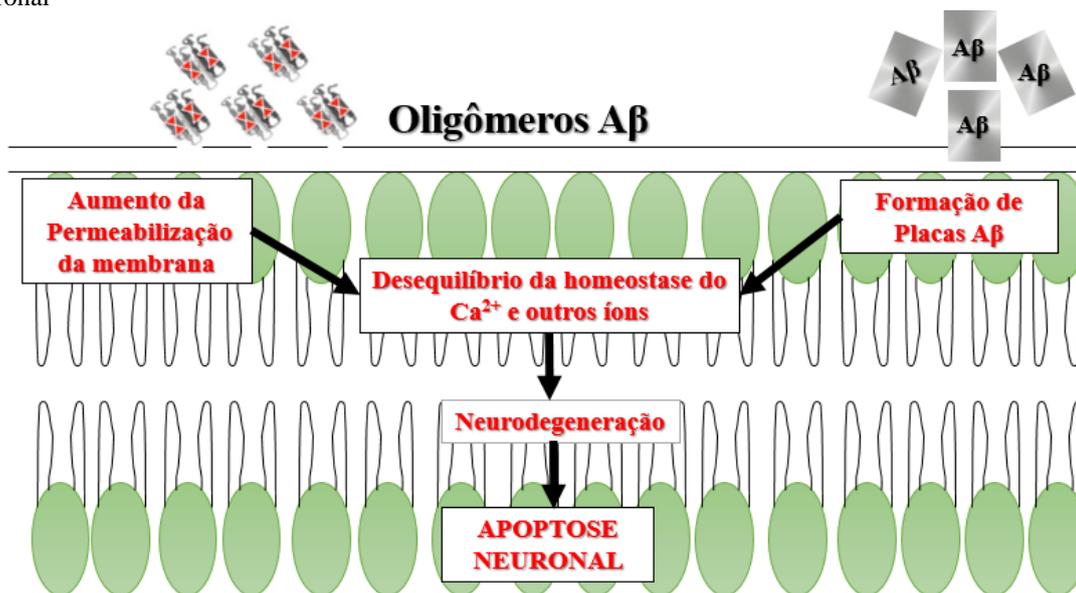
Os FHPs possuem de seis a oito moléculas de fosfatos no entorno da proteína tau (estado de hiperfosforilação), ao passo que nos cérebros saudáveis, tem-se por volta de dois grupos. O estado de hiperfosforilação acarreta congestionamento nesta proteína, resultando no impedimento do transporte de informações axonal até o dendrítico. Esse impedimento

compromete toda a arquitetura celular e provoca apoptose neuronal, característica preponderante em pacientes que apresentam distúrbios neurodegenerativos.<sup>18,19</sup>

Já a hipótese da cascata amiloide é vista como um modelo quantitativo linear centrado no neurônio e pode estar diretamente ligada a uma série de acontecimentos negativos, como consequência da deposição de placas  $\beta$ A (intra e extracelular), por exemplo, avanço da neuropatologia em função da proteína tau, disfunções sinápticas, inflamações, perdas neuronais.<sup>20</sup>

Não obstante, de acordo com os estudos de Benilova e colaboradores<sup>21</sup>, essa linearidade do modelo ainda é vaga e controversa, pois a ligação direta entre a deposição das placas  $\beta$ -amiloides e a toxicidade do neurônio resulta em achados confusos com a participação de pelo menos dez mecanismos celulares e vários receptores. A seguir (Figura 2), é apresentado um dos principais mecanismos.

**Figura 2-** Os oligômeros  $A\beta$  se agregam formando placas  $\beta$ -amiloides. Estas, por sua vez, entram na membrana e abrem os canais de  $Ca^{2+}$  e de outros íons, causando excitabilidade nas células neuronais e, posteriormente, apoptose neuronal



**Fonte:** autoria própria

O peptídeo  $\beta$ A possui de 39-42 aminoácidos em sua formação, resultado do metabolismo da proteína precursora de amiloide, APP (do inglês, *Amyloid Precursor Protein*), cujo papel é de grande importância nos processos de neuroplasticidade.<sup>22,23</sup> A APP é clivada por duas proteínas, a  $\gamma$ -secretase e a  $\beta$ -secretase, gerando dois fragmentos de peptídeos de tamanhos distintos, o  $\beta A_{1-40}$  e o  $\beta A_{1-42}$ . Acredita-se que o segundo peptídeo, embora menos comum que o primeiro e mais hidrofóbico, seja mais amiloidogênico. No entanto, ambos são os responsáveis pela

deposição de protofibrilas com estrutura intermediária entre os NFTs (placas  $\beta$ A, fibrilas) e as placas insolúveis em células neuronais.<sup>24</sup>

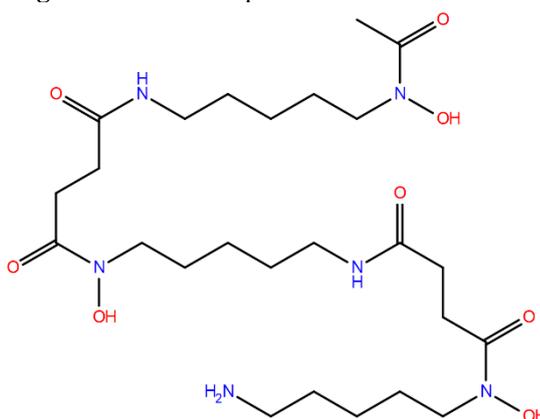
Em uma outra linha, estudos mostram que metais fisiológicos como Zn, Cu e Fe interagem com as placas amiloides.<sup>25</sup> Essa interação promove a catálise de radicais hidroxilas, acarretando em estresse oxidativo e o resultado desse evento é o surgimento de fibras amiloides tóxicas, as quais possuem grande predisposição à agregação e resistência para serem removidas.<sup>26-30</sup>

Esses achados permitiram que pesquisadores analisassem possíveis utilidades de agentes quelantes de metais para o tratamento da doença de Alzheimer. No entanto, uma das dificuldades encontradas surgem do fato que esses compostos não atravessam a barreira hematoencefálica e nem apresentam especificidade para o complexo íon metal- $\beta$ A, resultando na quelação de vários outros metais com posterior efeitos colaterais, por causa do uso prolongado dessas substâncias.<sup>31,32</sup>

Um desses agentes quelantes é a desferroxiamina- DFO (Figura 3), fármaco aprovado pelo FDA (do inglês, *Food and Drug Administration*) para o tratamento da hemocromatose, doença associada ao excesso de ferro. Entretanto, os resultados a partir dos ensaios clínicos mostraram eficácia em inibir o avanço da DA, devido à quelação e remoção do ferro no cérebro.<sup>33,34</sup> Outra característica da DFO é sua habilidade em se complexar com cobre e zinco.<sup>35,36</sup>

Apesar dessas características positivas, a DFO não consegue atravessar a barreira hematoencefálica devido ao seu peso molecular elevado. Entretanto, os resultados mencionados foram obtidos via administração subcutânea prolongada.<sup>37,38</sup>

**Figura 3** - Estrutura química da desferroxiamina



**Fonte:** autoria própria

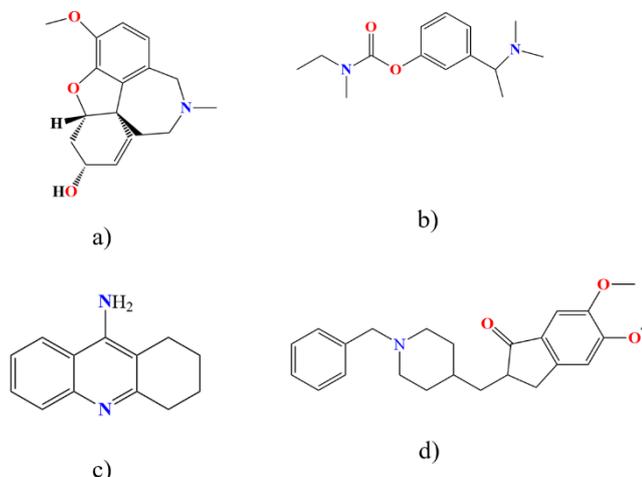
Portanto, diante de tudo que foi colocado, pode-se supor que essa hipótese fornece ingredientes para tentar explicar essa neuropatologia. Contudo, a linearidade do modelo, tendo o neurônio como ponto central, deve ainda ser melhor investigada para elucidar a participação

de todos os tipos de mecanismos celulares bem como a interação entre ambos (caso esta exista), provendo, dessa forma, aporte para explicar a evolução progressiva da DA. No que se refere a potenciais fármacos que poderiam atuar nesses tipos de alvos moleculares, muito esforço deverá ser empregado para desenvolver medicamentos que sejam capazes de remover os íons metálicos e, ao mesmo tempo, consigam atravessar a barreira hematoencefálica.

### 1.2 Hipótese Colinérgica

A hipótese colinérgica é uma das hipóteses mais antiga e mais bem aceita que tenta explicar a doença de Alzheimer. Essa hipótese eclodiu nos anos 1980, publicada por Bartus e colaboradores<sup>38</sup> e foi a base para o desenvolvimento dos fármacos em uso, por exemplo galantamina, rivastigmina, tacrina (atualmente removida do mercado devido a efeitos colaterais severos) e donepezil (Figura 4).

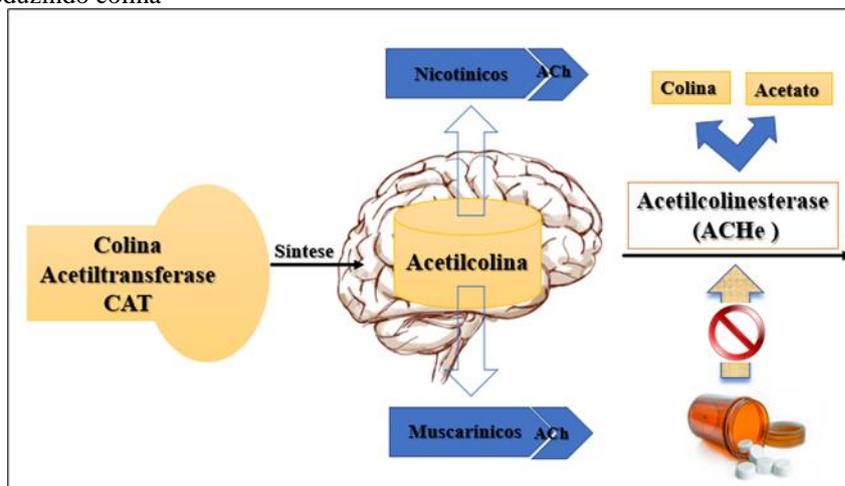
**Figura 4-** Fármacos usados no tratamento da DA. **a)** galantamina, **b)** rivastigmina, **c)** tacrina, **d)** donepezil



**Fonte:** autoria própria

A característica principal da associação do sistema colinérgico com a DA é a redução na concentração da colina acetiltransferase (CAT), enzima responsável pela síntese da acetilcolina (ACh) no córtex e no hipocampo. Uma vez produzida, parte da ACh é liberada na fenda sináptica, onde poderá ser acoplada a dois tipos de receptores, muscarínicos e nicotínicos. A outra parte é degradada pela enzima acetilcolinesterase (ACHE) em colina e acetato (Figura 5). Os fármacos (Figura 4) que atuam nesse sistema agem como inibidores da degradação da acetilcolina.<sup>38,39</sup>

**Figura 5**-Mecanismo da hipótese colinérgica- a colina acetiltransferase sintetiza acetilcolina que, por sua vez, se liga a dois receptores (nicotínicos e muscarínicos). Por meio da ação da enzima acetilcolinesterase, a acetilcolina é degradada, produzindo colina



Fonte: autoria própria

Os fármacos (Figura 4), inibidores da acetilcolinesterase, são administrados com o intuito de impedir a degradação da acetilcolina. Logo, possuem a função de reduzir o declínio cognitivo, impulsionando os processos de aprendizagem e memória.

Ainda que os inibidores da acetilcolinesterase sejam os fármacos prescritos a pacientes acometidos pela DA, essa classe de fármacos apresenta algumas limitações, a saber: efeitos colaterais (náusea, diarreia e tremores) e eficácia identificada apenas nos pacientes que se encontram nos estágios leve a moderado da doença.<sup>40</sup> Além do mais, a hipótese colinérgica não é suficiente para explicar os mecanismos que desencadeiam essa neuropatologia. Pois, há evidências que não só a redução da acetilcolina está associada ao Alzheimer, outras hipóteses envolvendo outros neurotransmissores também estão.<sup>41</sup>

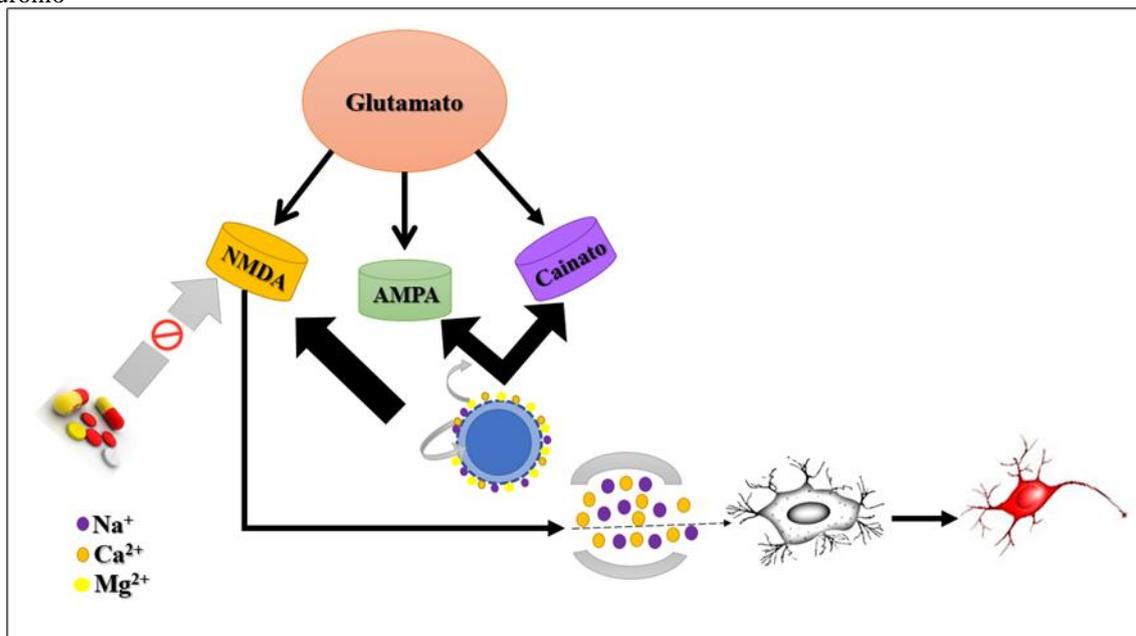
### 1.3 Hipótese glutamatérgica

Em meados dos anos 1980, a hipótese glutamatérgica veio à tona. O glutamato, principal neurotransmissor excitatório do sistema nervoso central, tem sua atividade mediada por três tipos de receptores, NMDA (*N-Metil-D-Aspartato*), AMPA ( $\alpha$ -amino-3-hidróxi-5-metil-4-isoxazolepropiónico) e Cainato.<sup>41,42</sup>

De acordo com essa hipótese, em estados específicos, por exemplo, alteração do metabolismo celular, os receptores NMDA são ativados excessivamente, podendo modificar a homeostase do cálcio ( $Ca^{2+}$ ) - demasiada liberação de  $Ca^{2+}$ . Como consequência, haverá uma

imensa excitabilidade do neurônio, resultando em apoptose neuronal (morte da célula neuronal), conforme visto na Figura 6.<sup>41,43</sup>

**Figura 6** - Mecanismo da hipótese glutamatérgica- o glutamato age por meio de três tipos de receptores, NMDA, AMPA e Cainato. Uma vez que os receptores NMDA são ativados de maneira exagerada, íons  $\text{Ca}^{2+}$  são liberados demasiadamente. Os receptores AMPA e Cainato também estimulam a entrada de íons  $\text{Ca}^{2+}$ , ao mesmo tempo que afastam os íons  $\text{Mg}^{2+}$ . Este por sua vez, atua bloqueando o entorno da membrana dentro do canal NMDA. Como resultado, o glutamato se liga a outra subunidade do receptor NMDA, atuando junto com a glicina que está ligada a outra subunidade. Toda essa cascata de evento resulta em mais liberação de íons  $\text{Ca}^{2+}$ , provocando a morte do neurônio



Fonte: autoria própria

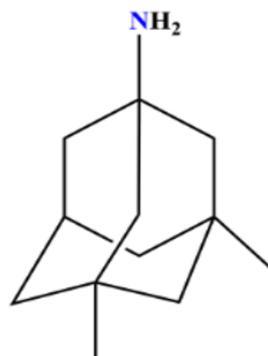
Os receptores NMDA são de difícil compreensão, uma vez que possuem vários sítios de ligação para o glutamato e inúmeros moduladores. Enquanto a afinidade dos receptores AMPA e Cainato pelo glutamato é considerada baixa, mas a interação do NMDA com esses dois receptores estimula a despolarização da célula mediante entrada de íons  $\text{Na}^+/\text{Ca}^{2+}$ , afastando, por sua vez, os íons  $\text{Mg}^{2+}$ , os quais atuam bloqueando o entorno da membrana dentro do canal do receptor NMDA.<sup>44</sup>

Esse evento possibilita ao glutamato ligar-se a uma subunidade do receptor NMDA, promovendo, em consonância com a glicina que está ligada a outra subunidade, a entrada de mais íons  $\text{Ca}^{2+}$  e  $\text{Na}^+$  nas células neuronais, provocando a excitabilidade do neurônio (Figura 6).<sup>44-46</sup> Acredita-se que essa cascata de eventos resulta na excitotoxicidade do glutamato, e esta, por sua vez, é um dos mecanismos fundamentais para que o processo neurodegenerativo seja gradualmente aumentado.<sup>47</sup>

O fármaco utilizado para bloquear o receptor NMDA, memantina (Figura 7), geralmente é administrado em estágios moderado e avançado da doença em combinação com outros

fármacos (inibidores colinérgicos)<sup>45</sup> e provavelmente atua na regulação do excesso da entrada de cálcio na célula neuronal, bloqueando os eventos de morte celular.

**Figura 7-** Estrutura química da memantina



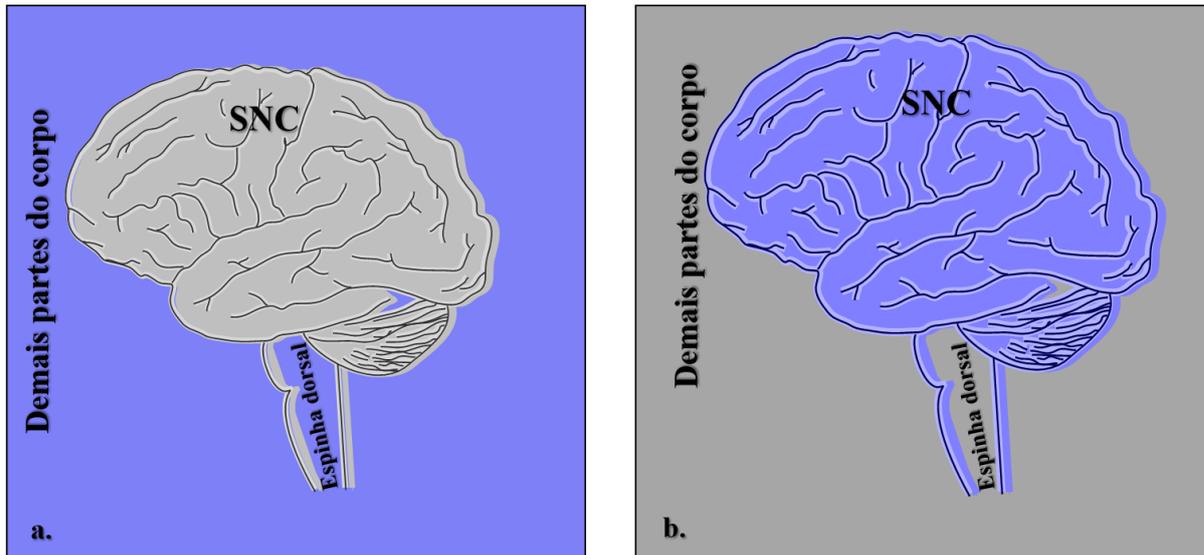
**Fonte:** autoria própria

Conforme abordado, as diferentes hipóteses divergem em relação às quais características fisiopatológicas mais contribuem para o desencadeamento e progressão da DA. Essas características estão atreladas a vários mecanismos, levando a abordagens terapêuticas diferentes. Porém, nenhuma é eficiente em fornecer aporte mecanístico claro, carecendo de mais estudos para planejar novos fármacos que atuem eficazmente no combate dessa enfermidade. Assomado a todos esses percalços, na busca por novas terapias medicamentosas, a barreira hematoencefálica se apresenta como um grande obstáculo a ser suplantado.

#### 1.4 A barreira hematoencefálica- um obstáculo a ser contornado

Nosso corpo apresenta várias barreiras, cuja função é nos proteger. Uma das mais conhecidas é a pele, e outra menos conhecida, mas de vital importância, é a barreira hematoencefálica (BHE). A BHE protege todo o sistema nervoso central (SNC) desde o cérebro, cerebelo até a medula espinhal contra a invasão de qualquer substância estranha. A BHE foi descoberta há mais de um século - pesquisadores introduziram corante azul na corrente sanguínea de um animal e perceberam que todos os tecidos coraram, exceto o cérebro. Em outros experimentos, ao injetarem o mesmo tipo de corante no SNC, perceberam-no azul, entretanto as outras partes do corpo permaneceram incólume (Figura 8). Na tentativa de explicar tais observações, os pesquisadores argumentaram que provavelmente haveria uma barreira que impedira o corante de atravessá-la, a barreira hematoencefálica.<sup>48-50</sup>

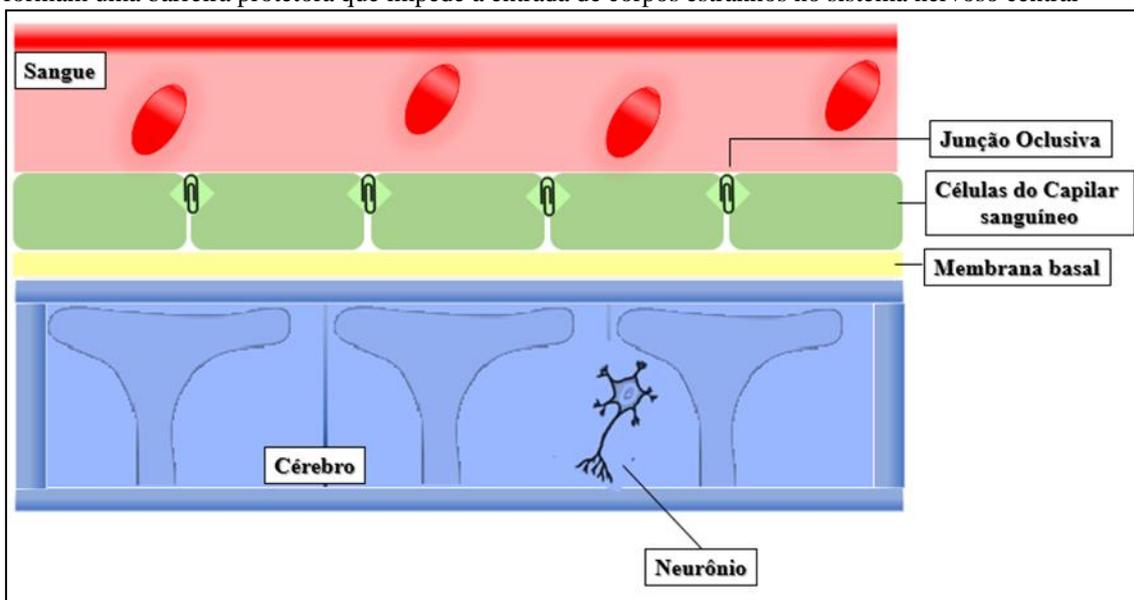
**Figura 8** - Representação dos experimentos que culminaram na descoberta da barreira hematoencefálica. **a)** Injeção do corante na corrente sanguínea; **b)** Injeção do corante no SNC



Fonte: autoria própria

A barreira hematoencefálica é composta por células de capilares sanguíneos e pela membrana basal. A BHE é uma membrana semipermeável, isto é, ela seleciona as substâncias que podem atravessá-la. Sua seletividade ocorre principalmente em virtude da configuração dos capilares sanguíneos e pela membrana basal de células da meninge. Estas células ficam extremamente próximas umas das outras, constituindo as junções oclusivas e impedindo que invasores externos penetrem a BHE.<sup>50</sup> A Figura 9 apresenta um diagrama esquemático da barreira hematoencefálica.

**Figura 9** - Diagrama esquemático da BHE - as junções oclusivas, células do capilar sanguíneo e membrana basal, formam uma barreira protetora que impede a entrada de corpos estranhos no sistema nervoso central



Fonte: autoria própria

As junções oclusivas são um complexo multiprotéico formado pelas junções apertadas (JA) e pelas junções endoteliais aderentes (JEA), sendo elas as principais reguladoras da permeabilidade celular.<sup>49,51</sup>

As JAs são constituídas por três proteínas integrais de membrana, claudina, ocludina, e moléculas de adesão juncional, além de proteínas citoplasmáticas acessórias. O papel das proteínas acessórias é conectar as proteínas de membrana à actina, a fim de preservar a estrutura e a funcionalidade do endotélio. Ao passo que as JEAs são formadas por proteínas de membranas denominada de caderinas que se ligam com a actina mediante proteínas intermediárias, chamada catedinas ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), formando ligações adesivas entre as células, e elas atuam em consonância com as JAs.<sup>50,52</sup>

No que concerne à doença de Alzheimer, estudos apontam fortes relações entre essa neuropatologia e a deterioração da barreira hematoencefálica. Pesquisas mostraram<sup>53</sup> que a deterioração da BHE tem forte dependência com o avanço da idade no hipocampo, região do cérebro associada à memória e à aprendizagem. Entretanto, em pacientes com declínio cognitivo moderado, esse fenômeno é mais brando, e a partir desse ponto os estágios iniciais dessa doença começam a serem perceptíveis.

Outros estudos indicam<sup>53-55</sup> que deposições de placas  $\beta$ -amiloides na vasculatura cerebral resultam em eventos inflamatórios e citotóxicos, o que acarreta em uma maior permeabilidade da BHE em indivíduos acometidos por DA. Estudos *in vitro* e modelos de patologia associados

à proteína tau em macacos transgênicos apontam<sup>56-59</sup> que essa proteína também participa do processo de deterioração da barreira hematoencefálica.

Portanto, com base nos estudos apresentados, pode-se inferir que tanto a deposição de placas  $\beta$ -amiloides quanto a taupatologia estão envolvidas no processo de perda da integridade da BHE, impulsionando os processos inflamatórios nas células neuronais e, por conseguinte, apoptose neuronal.

Diante do que foi apresentado, acredita-se que o problema da falta de fármacos que atuem efetivamente no combate às doenças neurodegenerativas está associado ao pouco conhecimento dos neurocientistas sobre a barreira hematoencefálica, conforme apontam alguns dados estatísticos:<sup>31,60</sup>

- Mais de 98% dos candidatos a fármacos formados por pequenas moléculas não atravessam a BHE<sup>31,60</sup>;
- Aproximadamente 100% de todos os fármacos constituídos por grandes moléculas, por exemplo, produtos de biotecnologia, não penetram a BHE<sup>31,60</sup>;
- Nenhuma grande empresa no mundo tem um protocolo de desenvolvimento de fármacos baseado no alvo para atuarem no SNC<sup>31,60</sup>;
- Se uma empresa desejar iniciar um programa de desenvolvimento de fármacos baseado no alvo para combater doenças relacionadas ao cérebro, enfrentará problemas acerca de recursos humanos. O quadro de profissionais com conhecimento a respeito da barreira hematoencefálica é escasso, pois poucos programas acadêmicos de neurociência focam na biologia do transporte da BHE, muito menos em alvos de drogas que atuam nela.<sup>31,60</sup>

A barreira hematoencefálica impõe alguns limites para que uma molécula possa transpô-la:<sup>31</sup>

- Peso molecular menor que 400 Daltons (Da);
- Número de ligação de hidrogênio abaixo da faixa 8-10.

O peso molecular é um dos assuntos mais discutidos quando se fala em desenvolver fármacos para atuarem no SNC. Pesquisas mostram<sup>61</sup> que abordagem com base na técnica THS

(do inglês, *High-Throughput Screening*) aumentam a seleção de fármacos com peso molecular elevado, sendo o limite do peso molecular uma propriedade atrelada à penetração nas membranas biológicas.<sup>62</sup> Um experimento usando a técnica THS mostrou que todas as drogas que agem no SNC possuem peso molecular maior que 426 Da, e a permeabilidade da BHE reduz quando a área de superfície de um fármaco é aumentada de 50 Å<sup>2</sup> para 100 Å<sup>2</sup>.<sup>63</sup>

Em relação ao papel das ligações de hidrogênio no processo de desenvolvimento de fármacos, estas devem ser consideradas concomitantes ao peso molecular. Os seguintes dados foram observados<sup>64</sup>:

- Inibidores da acetilcolinesterase (ACHE) formam de 3 a 5 ligações de hidrogênio;
- De modo geral, os fármacos utilizados no SNC têm de 1 a 5 ligações de hidrogênio;
- Metade dos inibidores  $\gamma$ -secretase usados no combate à doença de Alzheimer têm de 8 a 10 ligações de hidrogênio, acarretando em restrições para penetrar a BHE;
- Inibidores  $\beta$ -secretase possuem de 8 a 12 ligações de hidrogênio, gerando dificuldade para adentrar a barreira hematoencefálica.

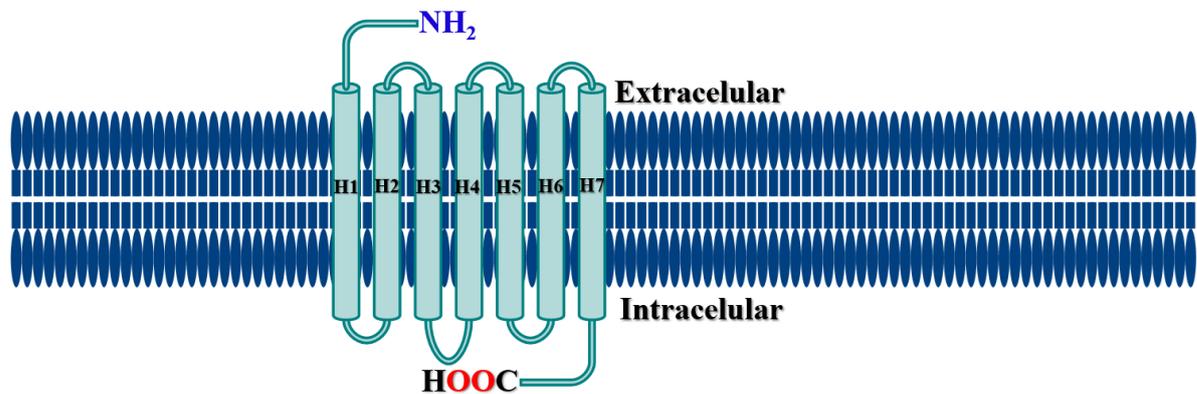
Diante do quadro apresentado, um dos principais entraves a ser superado, quando se pretende desenvolver fármacos para atuarem no SNC, é a barreira hematoencefálica. Portanto, deve-se planejar candidatos a fármacos não apenas focando no alvo biológico, mas também nos obstáculos encontrados diante da BHE. No que se refere à doença de Alzheimer, ante as limitações dos fármacos administrados aos pacientes acometidos por essa neuropatologia, faz-se necessárias pesquisas abordando novas terapias para combatê-la. Nesse cenário, um novo e promissor alvo surgiu, o receptor 5-HT<sub>6</sub> (5-hidroxitriptamina, subtipo 6).

As pesquisas têm mostrado<sup>40,65,66</sup> que esse receptor está intrinsecamente relacionado ao desencadeamento da doença de Alzheimer e seu bloqueio melhora significativamente os processos de aprendizagem e memória, os quais são bastante prejudicados em indivíduos acometidos por essa neuropatologia.

### 1.5 Receptor 5-HT<sub>6</sub>

O receptor 5-HT<sub>6</sub> pertence à família dos receptores da serotonina (5-HT) e foi um dos mais recentes a ser identificado, fruto do trabalho de três grupos de pesquisa independentes em 1993.<sup>67,68</sup> Em conformidade como os demais receptores do seu grupo, com exceção do receptor 5-HT<sub>3</sub> (associado ao canal iônico), o 5-HT<sub>6</sub>R é membro da grande família GPCRs (do inglês, *G-Protein Coupled Receptor*). Os receptores dessa família possuem três *loops* extracelulares (ECLs) e três intracelulares (ICLs), sendo receptores transmembranares (7TM), isto é, transpassam a membrana várias vezes<sup>69,70</sup> (Figura 10).

Figura 10- Estrutura geral de um receptor 7TM.



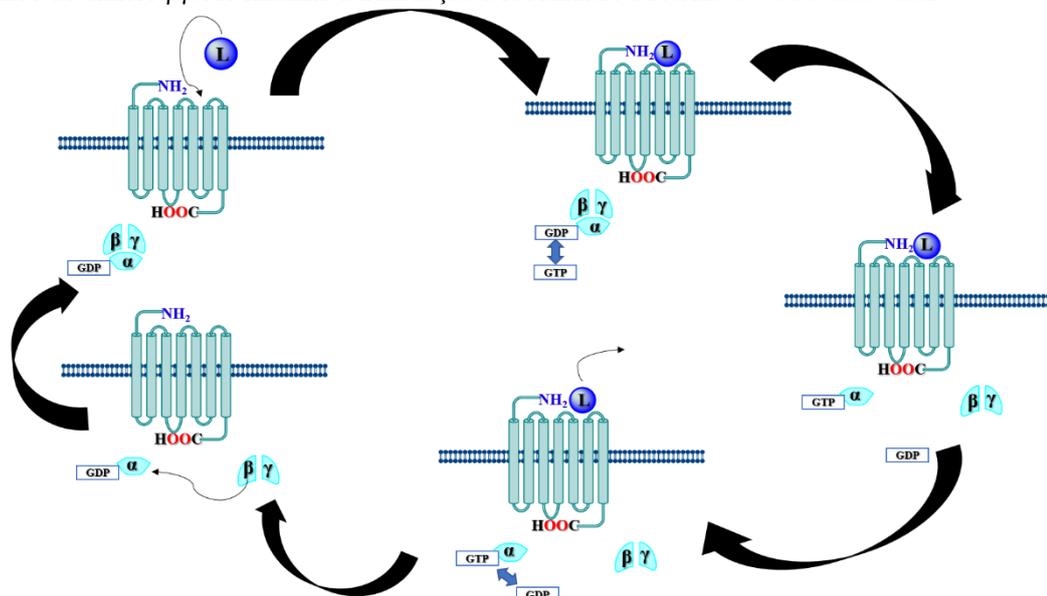
Fonte: autoria própria

Os receptores acoplados à proteína G fazem parte da maior e mais importante família de receptores, os quais respondem a uma ampla variedade de estímulos fisiológicos e ambientais. Vários GPCRs são importantes alvos terapêuticos para o tratamento de diversas enfermidades, por exemplo, câncer, problemas cardíacos, doenças associadas ao sistema nervoso central, problemas inflamatórios, desequilíbrio metabólico, entre outras. Tudo isso coloca os GPCRs como alvos de aproximadamente um terço dos fármacos atualmente comercializados, e por isso são focos de pesquisas na busca de mais abordagens terapêuticas para o tratamento de diferentes patologias, principalmente aquelas relacionadas ao sistema nervoso central.<sup>70</sup>

A atuação de um GPCR, por exemplo, o 5-HT<sub>6</sub>R, ocorre da seguinte forma: O N-terminal, localizado na parte extracelular e os ECLs são responsáveis pelo reconhecimento e modulação dos ligantes. Nesta etapa, as hélices transmembranares H5 e H6 se deslocam para o exterior, formando uma fenda no receptor. Posteriormente, o C-terminal e a parte ICLs

interagem com a proteína G, ativando-a, e, por conseguinte, desencadeando vários eventos celulares.<sup>70</sup> A proteína G, por sua vez, é composta pela união de três subunidades,  $\alpha$ ,  $\beta$  e  $\gamma$ . A subunidade  $\alpha$  que atua no reconhecimento dos receptores está ligada à guanosina difosfato inativa (GDP), formando no estado basal o complexo receptor  $\alpha$ -GDP. Quando o receptor interage com seu ligante, ocorre a troca de GDP por guanosina trifosfato ativa (GTP) e a separação da subunidade  $\alpha$  do dímero  $\beta\gamma$ , etapa seguida pelo desencadeamento de diversos eventos celulares. A proteína G continua ativa até que a subunidade  $\alpha$  retorne ao seu estado basal, religando-se ao GDP e reestabelecendo a interação com o dímero  $\beta\gamma$ , finalizando a cascata de eventos (Figura 11).<sup>70</sup>

**Figura 11-** Mecanismo geral de interação dos GPCRs com as subunidades da proteína G - na etapa inicial, o receptor interage com o ligante e ativa a proteína G. A subunidade  $\alpha$  se desliga da GDP para se ligar à GTP, dissociando do dímero  $\beta\gamma$ . Ao finalizar a transdução a subunidade  $\alpha$  retorna ao seu estado basal.



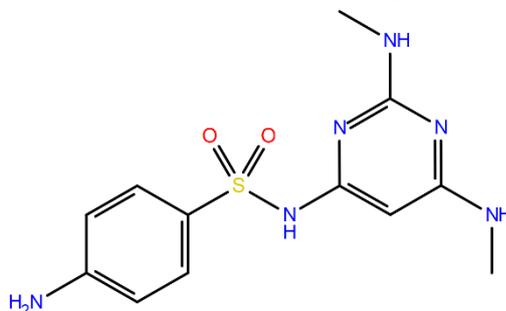
Fonte: autoria própria

Um dos GPCRs que se tornou alvo de interesse nos últimos anos foi o receptor 5-HT<sub>6</sub>. Devido a sua localização em áreas do cérebro envolvendo os processos de memória e aprendizagem, o 5-HT<sub>6</sub> configurou-se como um potencial alvo terapêutico ante a doença de Alzheimer.<sup>67,68</sup> Desde a sua identificação, tremendos esforços foram realizados para tentar elucidar seu papel funcional.<sup>71,72</sup> E um desses estudos foi utilizando Oligonucleotídeo Anti-sentido (OA), sistema artificial cuja função é bloquear a transdução do mensageiro no citoplasma quando absorvido na célula; ou no núcleo, impedindo o processamento nuclear; ou ainda em sistemas virais impedindo a replicação do vírus. Os estudos com esse sistema demonstraram mudanças comportamentais que podem ser bloqueadas por antagonistas

muscarínicos, indicando um papel modulatório do receptor 5-HT<sub>6</sub> na função colinérgica.<sup>40</sup> Esses estudos também mostraram o aumento do desempenho de roedores diante da realização de tarefas no labirinto aquático de Morris, sugerindo o envolvimento desse receptor em processos cognitivos.<sup>72</sup>

A influência do receptor 5-HT<sub>6</sub> no sistema colinérgico foi confirmada posteriormente por meio da administração do antagonista Ro 04-6790 (Figura 12) em roedores. Uma vez que o receptor 5-HT<sub>6</sub> foi bloqueado, vários eventos neurofarmacológicos foram observados, por exemplo, bocejar, mastigar, alongar-se, locomover-se, além de aquisição e consolidação de memória, processos esses associados à liberação de acetilcolina no hipocampo.<sup>73</sup>

**Figura 12-** Estrutura química do Ro 04-6790 - antagonista do receptor 5-HT<sub>6</sub>.



**Fonte:** autoria própria

Sabe-se que ao se administrar um antagonista muscarínico, todos os eventos neurocelulares anteriormente mencionados podem ser suprimidos.<sup>73</sup> De fato, a administração concomitante de atropina (antagonista muscarínico) e o antagonista Ro 04-6790 (Figura 12) resultou na supressão de bocejo. O mesmo resultado foi obtido substituindo a atropina pela escopolamina (outro antagonista muscarínico).<sup>74</sup>

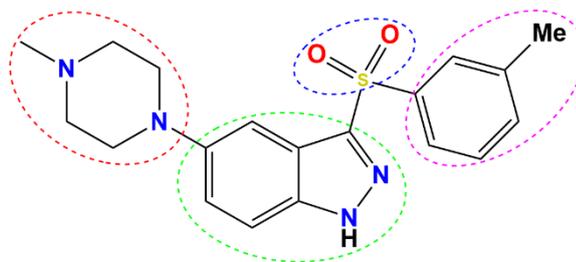
Vários outros estudos também evidenciaram que a função do receptor 5-HT<sub>6</sub> poderia estar relacionado à modulação do sistema colinérgico.<sup>74-78</sup> Esses grupos de pesquisas estudaram os efeitos do bloqueio desse receptor e a relação com a acetilcolina e mostraram que a administração de antagonistas do 5-HT<sub>6</sub>R, após uma tarefa de aprendizagem, revertem os efeitos amnésicos. Outras pesquisas apontaram<sup>79-81</sup> que além da interação do receptor 5-HT<sub>6</sub> com o sistema colinérgico, o seu bloqueio também aumenta a neurotransmissão no sistema glutamatérgico.

Evidenciada a importância do receptor 5-HT<sub>6</sub>, entende-se o motivo pelo qual ele se tornou um alvo de interesse na busca por novos antagonistas anti-amnésicos. Para planejar novos fármacos, é necessário conhecer algumas características intrínseca dos antagonistas desse receptor<sup>82</sup>, a saber (Figura 13):

- Átomo ionizável positivo (PI);
- Grupo aceptor de hidrogênio (HBA);
- região hidrofóbico (HYD);
- região hidrofóbico aromático (AR).

Geralmente, o grupo **PI** é representado por um anel piperazina; **HBA** é simbolizado por um grupo sulfonamida ou sulfonil; **HYD** é constituído por vários anéis aromáticos fundidos que geralmente estão ligados a halogênios, benzotiofenos ou índoles e o grupo **AR** é formado por anéis aromáticos heterocíclicos, tais como piridina, benzimidazoles ou quinolona.

**Figura 13** - Características dos elementos estruturais dos antagonistas do 5-HT<sub>6</sub>R. Vermelho: **PI**, Verde: **AR**; Azul: **HBA**; Magenta: **HYD**



Fonte: autoria própria

Na busca por novos antagonistas anti-amnésicos, tendo o receptor 5-HT<sub>6</sub> como alvo, esta tese de doutorado está centrada no planejamento de novos candidatos a antagonistas desse receptor, usando ferramentas computacionais.

No processo de desenvolvimento de um novo fármaco e até mesmo no reposicionamento de medicamentos, o uso de ferramentas computacionais tornou-se de extrema importância. A aplicação delas otimiza a busca por novos agentes terapêuticos, reduzindo o tempo e os custos associados, além de aumentar a chance de sucesso. Por isso, tanto as indústrias como as academias estão empregando-as em suas pesquisas para que assim consigam colocar no mercado abordagens terapêuticas em tempo reduzido, sem deixar de lado sua eficácia e segurança. Para isso, várias técnicas são utilizadas, tais como, relações entre a atividade biológica e a estrutura química (SAR), relações quantitativas entre atividade biológica e a estrutura química (do inglês, *Quantitative Structure-Activity Relationships*, QSAR 2D, 3D, 4D), ancoramento molecular (*docking* molecular), *Dinâmica* molecular e modelagem por homologia (predição das coordenadas tridimensionais de uma proteína).<sup>83</sup>

## 2 APLICAÇÃO DE MÉTODOS COMPUTACIONAIS NO PLANEJAMENTO DE FÁRMACOS

Duas são as abordagens quando se deseja desenvolver fármacos - Abordagem focada no receptor (do inglês, *Structure-Based drug design* - SBDD) e abordagem focada no ligante (do inglês, *Ligand-Based Drug Design*- LBDD), quando não se conhece a estrutura do receptor.<sup>84,85</sup>

### 2.1 Abordagem molecular focada no receptor

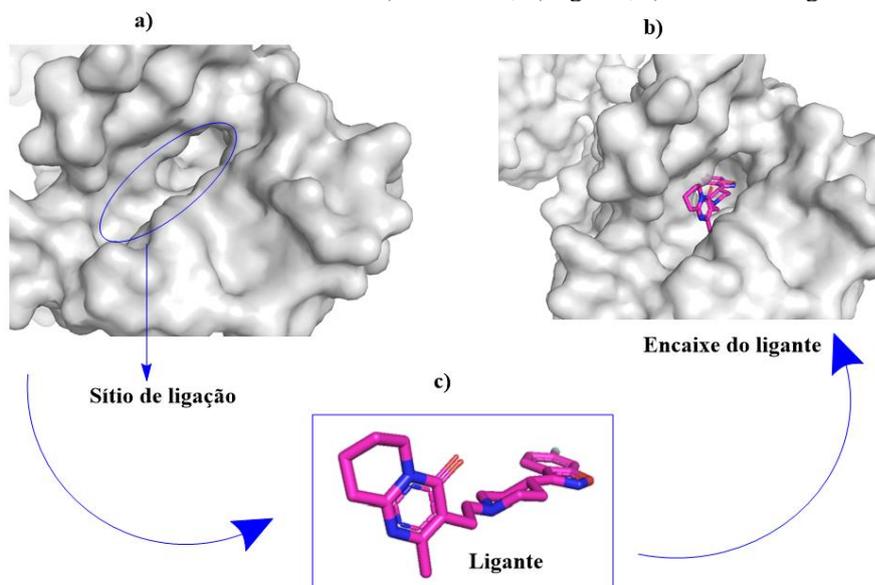
O planejamento baseado na estrutura do receptor (SBDD) lança mão das informações extraídas do arranjo topológico dos alvos moleculares (proteína, enzimas, DNA, dentre outros) para propor novos compostos. Essas informações são oriundas das estruturas tridimensionais (3D) obtidas via técnicas de raios X, estudos de RMN (Ressonância Magnética Nuclear), microscopia eletrônica ou modelagem por homologia (quando não há estrutura 3D determinada experimentalmente).<sup>86, 87</sup>

#### 2.1.1 Docking Molecular

O *Docking* Molecular é a principal abordagem SBDD utilizada no planejamento de fármacos. Por meio dessa metodologia é possível antever o processo de reconhecimento molecular, bem como o mecanismo que rege a interação receptor- ligante. O ancoramento molecular é foco de interesse tanto das universidades quanto da indústria farmacêutica, uma vez que o uso dessa técnica pode reduzir consideravelmente os custos associados à fase experimental.<sup>86</sup>

Conhecendo-se o sítio de ligação de um determinado alvo biológico, encaixa-se o ligante no sítio, de modo que cada grupo do ligante é pareado com o grupo complementar no sítio ativo, e então o programa busca pelo melhor ajuste do ligante dentro desse sítio<sup>88</sup>(Figura 14).

**Figura 14** - Processo de ancoramento molecular- a) sítio ativo; b) ligante; c) encaixe do ligante no sítio ativo



**Fonte:** autoria própria

A energia do complexo receptor-ligante é estimada por meio das poses de ligação mais prováveis. São duas as etapas empregadas nesse processo:

- Exploração de um amplo espaço conformacional, alcançando-se vários modos de ligação, uma vez que no processo de reconhecimento molecular um grande número de graus de liberdade é obtido (referentes às mudanças conformacionais do ligante, receptor e solvente envolvidos nesse processo);
- Previsão precisa da energia de interação de cada um dos modos de ligação.

Os vários softwares focados em *Docking* molecular executam essas etapas por meio de um algoritmo de busca, e os resultados são avaliados por uma função *scoring* (uma forma de avaliar a energia livre de ligação).

No processo de busca conformacional, os parâmetros estruturais dos ligantes (flexibilização do ligante), tais como graus de liberdade torcional (diedral), translacional e rotacional são modificados, e os programas usam diferentes métodos para realizar essa tarefa. Destacam-se dois:

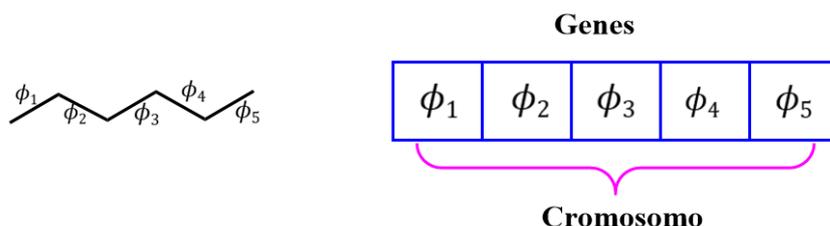
**Método de busca sistemática:** nesse método, variações graduais nas conformações do ligante são realizadas. Posteriormente, o algoritmo investiga o cenário energético dos espaços

conformacionais e após vários ciclos de busca e avaliação, a solução mínima de energia é encontrada, correspondente ao modo de ligação mais provável. A desvantagem desse método é que nem sempre a energia é convergida para um mínimo global. Na maioria das vezes, a energia é convergida para um mínimo local. Essa desvantagem é superada dividindo o ligante em partes e buscando pela melhor conformação para cada parte. Primeiramente, o fragmento principal (denominado de **Anchor**) é docado no sítio de ligação, e as demais partes são adicionadas, posteriormente. Esse tipo de método é incrementado em vários softwares, como DOCK, FlexX, Glide, EUDOCK e Surflex.<sup>89</sup>

**Método de busca estocástico:** esse tipo de método realiza busca conformacional mediante modificações aleatórias nos parâmetros estruturais do ligante. Para isso, o algoritmo gera vários *ensembles* com conformações moleculares aleatórias, abrangendo uma ampla cobertura do cenário energético. Isso impede que a solução final esteja atrelada ao mínimo local e aumenta a probabilidade de encontrar um mínimo global. Vários algoritmos são implementados para auxiliar nessa busca energética, como Monte Carlo, recozimento simulado (*Simulated Annealing*) e algoritmos evolucionários (algoritmo genético).

Nos métodos estocásticos, o algoritmo genético (AG) é o principal método implementado nos programas para fazer a busca conformacional (os demais algoritmos, *Simualted Annealing* e Monte Carlo são detalhados por Guedes e colaboradores.<sup>89</sup> Inicialmente, o método AG codifica todos os parâmetros estruturais da estrutura molecular (representando diferentes conformações, cujos valores dos ângulos torcionais são atribuídos aleatoriamente). Cada conformação corresponde a um cromossomo, e cada valor de ângulo torcional representa um gene (Figura 15). A partir desse cromossomo inicial, o algoritmo de busca gera uma população inicial de cromossomos, cobrindo uma ampla área do cenário energético. Essa população é avaliada e os cromossomos mais adaptados (aqueles com os melhores valores de energia) são selecionados para a geração da próxima população. Esse procedimento diminui a energia média do conjunto cromossômico e transmite as características estruturais favoráveis de uma população para outra, reduzindo o espaço conformacional a ser explorado até obter a conformação mais estável.

**Figura 15** - Representação do funcionamento de um AG - atribui-se aleatoriamente valores a cada ângulo torcional (gene), sendo o conjunto desses valores representado por um cromossomo.



Fonte: autoria própria

A predição da afinidade do ligante pelo sítio ativo do alvo biológico é realizada recorrendo-se a funções de pontuação (*scoring*). A variação energética é dada pela constante de ligação (Kd) e pela energia livre de Gibbs (G). A avaliação da energia é dada em função dos parâmetros físico-químicos, tais como interações moleculares, dessolvatação e efeitos entrópicos. Dessa forma, quanto maior o número de parâmetros avaliados, maior será o custo computacional. As funções *scoring* são categorizadas em:

**Funções baseadas em campos de força:** esse tipo de função estima a energia de ligação por meio da soma dos termos ligados (estiramento das ligações, ângulos de ligação e variação dos diedros) e dos termos não ligados (interações eletrostáticas e de Van der Waals), transformando-as em uma função geral. Essas funções usam métodos *ab initio* para calcular a energia associada a cada termo por meio de equações da mecânica clássica. A maior lacuna deste tipo de função é a falta de avaliação para termos entrópicos, devido à ausência do solvente, o qual não é contabilizado. Logo, não há como estimar a energia de dessolvatação.

**Funções do tipo empírica:** cada termo da função descreve um tipo de evento físico-químico envolvido na formação do complexo receptor-ligante. Estes termos incluem ligação de hidrogênio, interações iônicas e apolares, efeitos de dessolvatação e efeitos entrópicos.

As funções empíricas são construídas usando as análises de regressão linear múltipla, modelo construído com uma série de complexos receptor-ligante com valores de afinidade de ligação conhecidos. Posteriormente, as constantes de pesos geradas pelo método estatístico são usadas como coeficiente para ajustar os termos da equação. O problema desse tipo de função é sua dependência da precisão de dados utilizados para desenvolver o modelo.

**Função baseada em conhecimento:** esse tipo de função usa pares de potenciais de energia extraídos do complexo receptor-ligante para obter a função geral. Estes potenciais são gerados,

considerando a frequência com a qual dois diferentes átomos são encontrados dentro de uma determinada distância no conjunto de dados estruturais. Os diferentes tipos de interações são observados e classificados de acordo com os pesos atribuídos (conforme sua frequência de ocorrência), e a função *score* final é o resultado da soma dessas interações individuais. Como esse tipo de função não depende da reprodução da afinidade de ligação (métodos empíricos) ou cálculos *ab initio* (funções baseadas em campo de força), há um equilíbrio entre precisão e velocidade.

Portanto, diante do conhecimento da técnica de *docking* molecular, bem como da experiência em operar os diferentes softwares existentes no mercado relacionados ao ancoramento molecular é possível extrair informações importantes para auxiliar na seleção de candidatos a fármacos, além de planejar compostos com as características químicas adequadas para atuar perante a um determinado alvo biológico.

### 2.1.2 *Dinâmica Molecular*

É imprescindível que se simule um sistema receptor-ligante levando em consideração a flexibilidade dos seus componentes, e assim prever assertivamente o modo de ligação dos candidatos a fármacos, bem como suas propriedades cinéticas e termodinâmicas. Para esse fim, lança-se mão da técnica *Dinâmica Molecular* (DM), que consiste na computação dos movimentos das partículas de um sistema molecular em função do tempo. Esta simulação ocorre recorrendo-se à aplicação da segunda lei de Newton,  $F = ma$ , que em notação diferencial podemos escrever (Eq.1):

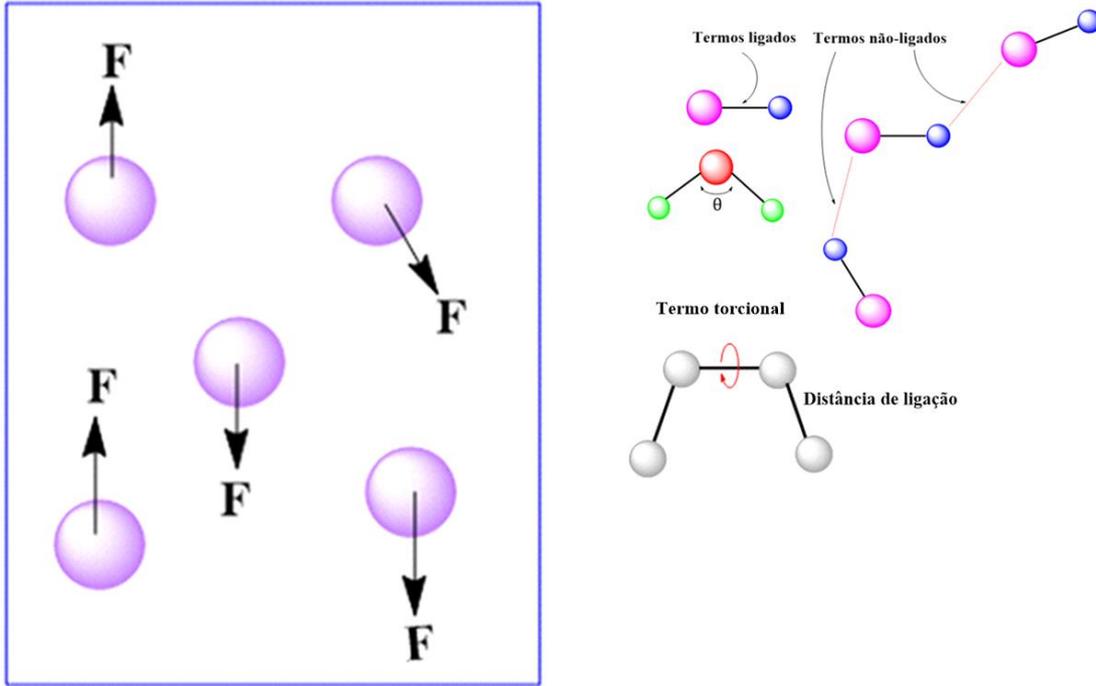
$$-\frac{dV}{dr} = \frac{d^2\mathbf{r}}{dt^2} m_i \quad (1)$$

Em que  $V$  é a energia potencial da partícula na posição  $r$ . O lado esquerdo da equação 1 é o negativo do gradiente da energia, denominado força ( $\mathbf{F}$ ) da partícula.<sup>90</sup>

As partículas em um sistema molecular interagem umas com as outras, logo, elas estão sujeitas a forças interatômicas e intermoleculares. Todas as forças que agem nas moléculas são descritas por uma função matemática denominado campo de força (descreve termos ligados como distância, ângulo de ligação e ângulos diédricos, além dos termos não-ligados, por exemplo, interações de Van der Waals e eletrostáticas das moléculas) conforme descrito na

Figura 16 e Eq.2. Existem vários campos de forças com suas vantagens e desvantagens, cabendo ao operador escolher aquele que melhor se enquadra ao problema que se deseja resolver.<sup>91</sup>

**Figura 16** - As partículas em um sistema molecular sofrem influência das forças de outras partículas. Em DM, todas as forças, termos ligados e termos não-ligados, são acopladas em uma função matemática denominada de campo de força.



Fonte: autoria própria

$$\begin{aligned}
 V = & \sum_i \frac{K_{l,i}}{2} (l_i - l_{0,i})^2 + \sum_i \frac{K_{\alpha,i}}{2} (\alpha_i - \alpha_{0,i})^2 \\
 & + \sum_i \left\{ \sum_k^M \frac{V_{i,k}}{2} [1 + \cos(n_{ik} \cdot \theta_{ik} - \theta_{0,ik})] \right\} \\
 & + \sum_{i,j} \varepsilon_{ij} \left[ \left( \frac{r_{0,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{0,ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_r r_{ij}}
 \end{aligned} \tag{2}$$

Os 3 primeiros termos da equação 2 descrevem as variações na energia potencial como uma função do comprimento de ligação, ângulo e torção entre os átomos diretamente envolvidos na ligação. Esses termos são representados pelo somatório do comprimento de ligação ( $l$ ), ângulos ( $\alpha$ ) e ângulos diedros ( $\theta$ ) dos átomos envolvidos na simulação.<sup>92</sup>

As contribuições de comprimento de ligação e ângulos possuem a mesma forma funcional, são descritos por potenciais harmônicos com valores de referência  $l_0$  e  $\alpha_0$ , constantes de força

Kl e Kw, respectivamente. No entanto, por causa da periodicidade intrínseca, os termos torcionais são naturalmente definidos por uma série de cosseno de M termos para cada ângulo diédrico. Assim,  $n_{ik}$  é um parâmetro que descreve a multiplicidade (quantidade de mínimos de energia) do K-ésimo termo da série,  $\theta_{0,ik}$  refere-se ao ângulo de fase correspondente, e  $V_{i,k}$  é a energia de barreira. Como todos esses parâmetros são atribuídos a átomos envolvidos diretamente em uma ligação química, essas contribuições são denominadas de termos ligados. O 4º e 5º termos na Eq.2 representam as interações de Van der Waals e eletrostáticas, respectivamente, entre os átomos, denominados de termos não-ligados.<sup>92</sup>

Escolhido o campo de força, para se analisar o movimento das partículas em função do tempo aplica-se uma determinada temperatura para alterar as posições dos átomos no espaço (o movimento das partículas deve ser o mínimo possível, de modo a considerar a aceleração quase constante), e, então, a posição e a velocidade de cada átomo do sistema em estudo são computadas. A partir desses resultados, retorna-se ao passo anterior e calcula a aceleração das partículas para descobrirmos as novas posições e velocidades no tempo seguinte. Esse cálculo é repetido até se obter trajetórias moleculares para todo o sistema em investigação, e para essa tarefa aplica-se a expansão de Taylor<sup>1</sup> (Eq. 3):

$$\begin{aligned} \mathbf{r}_{i+1} &= \mathbf{r}_i + \frac{\partial \mathbf{r}}{\partial t}(\Delta t) + \frac{1}{2} \frac{\partial^2 \mathbf{r}}{\partial t^2}(\Delta t)^2 + \frac{1}{6} \frac{\partial^3 \mathbf{r}}{\partial t^3}(\Delta t)^3 + \dots \\ \mathbf{r}_{i+1} &= \mathbf{r}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2} \mathbf{a}_1(\Delta t)^2 + \frac{1}{6} \mathbf{b}_i(\Delta t)^3 + \dots \end{aligned} \quad (3)$$

As velocidades  $\mathbf{v}_i$  são alcançadas empregando-se a primeira derivada das posições em função do tempo ( $\partial \mathbf{r} / \partial t$ ), a aceleração,  $\mathbf{a}_1$ , é o resultado da segunda derivada em função do tempo ( $\partial^2 \mathbf{r} / \partial t^2$ ), a hiperaceleração,  $\mathbf{b}_i$ , é a solução da terceira derivada, e assim por diante. As posições em um intervalo de tempo anterior são obtidas a partir da aplicação da Eq. 3 substituindo  $\Delta t$  por  $-\Delta t$  (Eq. 4).

$$\mathbf{r}_{i-1} = \mathbf{r}_i - \mathbf{v}_i(\Delta t) + \frac{1}{2} \mathbf{a}_1(\Delta t)^2 - \frac{1}{6} \mathbf{b}_i(\Delta t)^3 + \dots \quad (4)$$

---

<sup>1</sup>A expansão de uma função  $f(x)$  pode ser determinada a partir da origem  $x=a$  com  $a \neq 0$  através da série de Taylor:  
 $f(x) = f(a) + f^{(1)}(a)(x-a) + f^{(2)} \frac{a}{2!}(x-a)^2 + \dots + f^n \frac{a}{n!}(x-a)^n$

Somando as equações 2 e 3, é possível prever a posição da partícula em um dado intervalo de tempo posterior a partir do tempo atual, da posição e da aceleração atual (a aceleração,  $\mathbf{a}_i$ , pode ser calculada a partir da força), conforme mostra a Eq. 5.

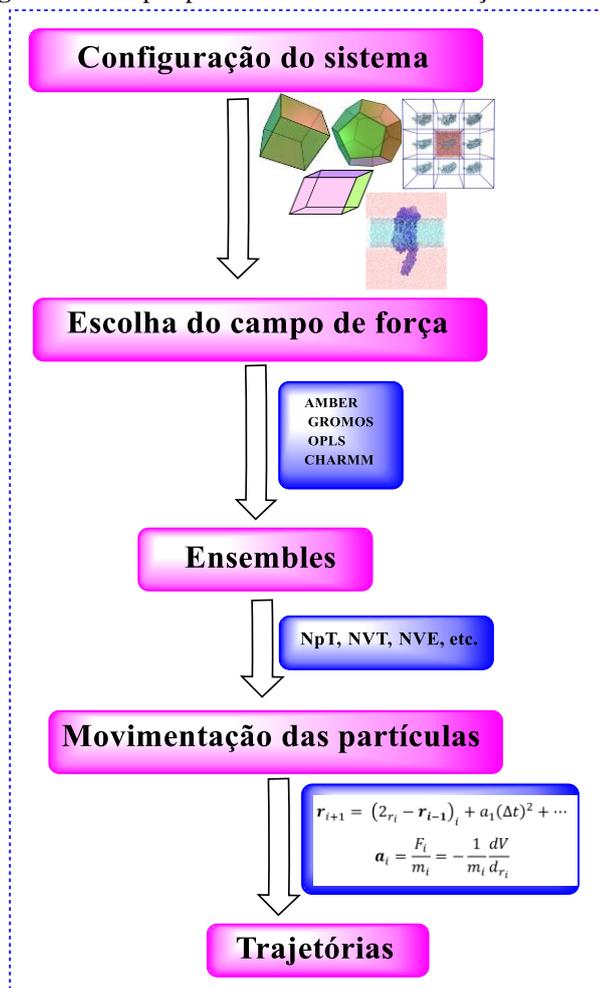
$$\mathbf{r}_{i+1} = (2\mathbf{r}_i - \mathbf{r}_{i-1})_i + \mathbf{a}_i(\Delta t)^2 + \dots$$

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} = -\frac{1}{m_i} \frac{dV}{d\mathbf{r}_i} \quad (5)$$

Este tipo de integração, dividida em pequenos intervalos de tempo ( $\Delta t$ ), é resolvida por meio da aplicação de alguns algoritmos. Em *Dinâmica* molecular, o algoritmo mais utilizado é o algoritmo de Verlet<sup>93</sup> (Eq.5). Nele, as posições e acelerações das partículas no tempo  $t$  e as do tempo anterior ( $t_{-1}$ ) são usadas para determinar as posições no tempo seguinte ( $t_{+1}$ ).

O processo de simulação de um sistema por meio do método de *Dinâmica* molecular percorre várias etapas, conforme se observa na Figura 17:

**Figura 17** - Etapas percorridas em uma simulação usando DM



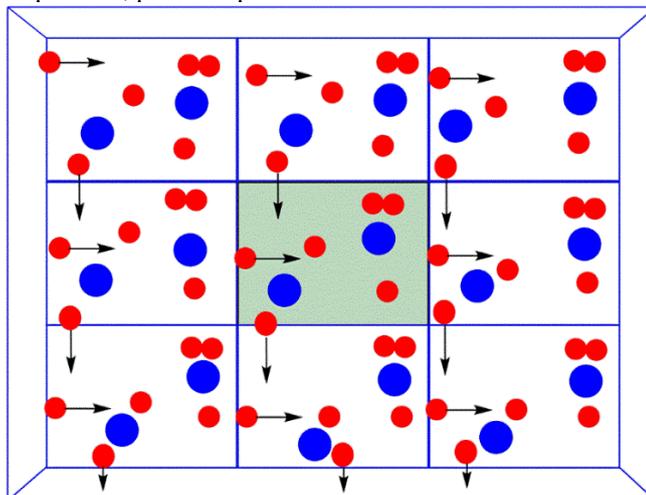
Fonte: autoria própria

## Configuração inicial do sistema

É impossível simular um sistema com o tamanho real, por isso a simulação é realizada com o menor número de partículas possível, de modo a reproduzir resultados semelhantes ao sistema real. Para isso, o procedimento geralmente adotado é “criar uma caixa de simulação” (espaço tridimensional em que todos os componentes do sistema são alocados), cujas dimensões se equiparem à densidade numérica ( $N/V$ ) do sistema real.<sup>91</sup>

Um grande problema surge ao adotar esse procedimento, supondo que tenhamos um sistema tridimensional com 1000 partículas, uma parte considerável dessas partículas está próxima às paredes interagindo com o vácuo, e a outra parte está no interior do sistema. No caso de proteínas, a nossa intenção é simular o ambiente fisiológico em que ela se encontra, e para isso é necessário impedir que essas partículas interajam com as paredes da caixa, ultrapassando-a. As condições periódicas de contorno são os artifícios empregados para contornar esse problema. Várias caixas são replicadas em todas as direções, de modo a manter a periodicidade do sistema. Dessa forma, apenas a caixa principal é simulada, e o movimento das partículas nas réplicas se iguala ao das partículas na caixa central (Figura 18). Dito de outra forma, se uma partícula se move para fora da caixa principal, uma outra partícula idêntica entrará pela direção oposta, e assim o número de partículas se conservará. Os efeitos das condições periódicas de contorno devem ser observados tanto nas equações de movimento quanto nos cálculos de interação. Após cada passo de integração, verifica-se se as coordenadas das partículas estão dentro da caixa, caso contrário suas coordenadas devem ser ajustadas, de modo que elas retornem à caixa.<sup>90</sup>

**Figura 18** - Condições periódicas de contorno - nesse tipo de simulação, apenas a caixa central é simulada. Embora o sistema seja delimitado, as partículas não interagem com as paredes da caixa, e assim consegue-se simular o ambiente fisiológico de uma proteína, por exemplo



Fonte: autoria própria

### Escolha do campo de força

A escolha do campo de força é difícil, pois há vários disponíveis na literatura, entre os quais podemos mencionar o GROMOS, OPLS, AMBER, CHARMM.<sup>94-97</sup> Cada um deles foi desenvolvido para reproduzir uma determinada propriedade de interesse. Então, é aconselhável conhecer bem o sistema em estudo para selecionar o campo de força mais apropriado aos interesses, de modo a obter resultados confiáveis.

Os campos de forças possuem suas particularidades e limitações, sendo impossível que um determinado campo de força reproduza todas as propriedades de um sistema molecular. Em se tratando de sistemas biológicos, como proteínas, a escolha é ainda mais complicada, pois os dados experimentais disponíveis são escassos. Diante dessa lacuna, as parametrizações são implementadas em um campo de força mediante cálculos *ab-initio*. Por exemplo, se quisermos conhecer os parâmetros de Lennard-Jones (Eq. 6) para um dado resíduo de aminoácido, primeiro montamos um sistema de simulação com o aminoácido escolhido adicionado a algumas camadas de hidratação, e depois realizamos o cálculo *ab-initio*. A partir dos resultados, ajustaremos os parâmetros de Lennard-Jones (o primeiro termo entre colchetes na equação 6 descreve as forças repulsivas, enquanto o segundo termo descreve as forças atrativas) de modo a reproduzir os mesmos resultados extraídos dos cálculos.<sup>91</sup>

$$V_{(r)} = 4 \epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (6)$$

## **Movimentação das partículas (equilíbrio)**

Anteriormente à simulação, o sistema sob investigação deve ser minimizado para que não haja maus contatos entre suas partículas e estas cheguem a um estado de equilíbrio com a menor energia potencial possível.

No caso de sistemas biológicos, como proteínas resolvidas via técnica de raios-X, quase sempre a concentração é superior à do ambiente fisiológico, e por isso, as partículas estão muito próximas umas das outras. Em outros casos, é necessário adicionar mais elementos para que o sistema se assemelhe ao estado do ambiente fisiológico (por exemplo, proteínas de membrana).<sup>90</sup> Portanto, a etapa de equilíbrio é o passo necessário para que o estado do sistema represente o objeto de interesse.

## ***Ensembles* e controle da simulação**

Nas simulações usando DM, configura-se o sistema para que algumas de suas propriedades sejam mantidas constantes. A escolha dessas configurações representa o estado do sistema denominado ensemble. Há vários tipos de *ensembles*, por exemplo, NpT, NVT, NVE ou  $\mu$ VT, em que N, p, T, E e  $\mu$  são número de partículas, pressão, temperatura, energia total do sistema e potencial químico, respectivamente.<sup>91</sup>

A partir do *ensemble* escolhido, várias informações são obtidas em cada tempo de integração (trajetória das partículas), gerando vários gigabytes de informações, o que torna impossível a análise de cada resultado. Por isso, no final de cada simulação o valor obtido para uma dada propriedade é, na verdade, a média de valores alcançados em cada passo da integração. Portanto, quanto maior o tempo de integração mais informações são geradas e maior a confiabilidade nos resultados.<sup>91</sup>

## **Análise e interpretação dos resultados**

A aplicação mais usual de *Dinâmica* molecular no campo de biologia molecular é a verificação da mobilidade ou flexibilidade de várias regiões de um dado alvo molecular, ou mesmo diante de estruturas, cujas coordenadas tridimensionais foram construídas via modelagem por homologia, a simulação pode ser realizada para refiná-la.<sup>91</sup> É possível, também, averiguar o comportamento de outros ligantes, selecionados a partir de banco de dados<sup>98-100</sup> ou

planejado para atuar no alvo de interesse, dentro do sítio ativo ao longo da simulação. Nessa mesma linha, pode-se avaliar o efeito funcional da mutação de alguns resíduos de aminoácidos, e a partir daí levantar hipóteses que posteriormente podem ser confirmadas ou refutadas via abordagens experimentais.

### 2.1.3 *Predição das coordenadas tridimensionais de uma proteína*

É imprescindível que se tenha as coordenadas tridimensionais do alvo biológico quando se deseja planejar fármacos empregando métodos SBDD. No entanto, ainda que haja uma vasta quantidade de estruturas 3D depositadas em banco de dados, tais como o PDB<sup>87</sup> (Do inglês, *Protein Data Bank*), nem todos os alvos biológicos são contemplados, uma vez que demanda muito tempo para se obter a estrutura final, recursos, e outras vezes, essa tarefa é tecnicamente difícil.

Na ausência de estruturas 3D determinadas experimentalmente, a comunidade científica lança mão de protocolos computacionais para prever as coordenadas tridimensionais dos seus alvos de interesses. Os métodos para predição de estruturas são basicamente divididos em duas categorias<sup>101</sup>: baseado em *template* (do inglês, *template-based*) e os independentes de *template* (do inglês, *template-free*).

No primeiro grupo, temos métodos de modelagem comparativa em que diante de uma sequência alvo, o algoritmo empregado busca por sequências evolutivamente relacionadas, cujas estruturas 3D já são conhecidas. Esses métodos baseiam-se na assertiva de que proteínas de uma mesma família partem de um mesmo ancestral, e, portanto, possuem semelhanças estruturais. Ainda nessa categoria, temos também o método de predição de enovelamento, em que embora as sequências de algumas proteínas apresentem baixa similaridade quando comparada a outras sequências, elas podem ser estruturalmente semelhantes. O resultado final para o alvo de interesse é obtido mediante a análise do tipo de enovelamento das estruturas-*templates*, em que as propriedades, tais como, estrutura secundária, polaridade das cadeias laterais e hidrofobicidade são avaliadas.<sup>101</sup>

Já na segunda categoria, temos os métodos independentes de moldes que incluem os métodos *ab initio* e de novo. Nessa categoria, as informações estruturais são obtidas via fragmentos de proteínas, estruturas secundárias e potenciais estatísticos extraídos de alvos, cujas sequências não apresentam similaridade com a sequência de interesse.<sup>101</sup>

Proteínas acopladas à proteína G estão entre os alvos mais complexos quando se deseja determinar suas estruturas tridimensionais. Como resultado, há um grande hiato entre o número de estruturas conhecidas e estruturas não resolvidas.<sup>98</sup> As proteínas GPCRs são responsáveis pela transdução de sinais nas membranas celulares e controlam diversos processos fisiológicos e biológicos. Por isso, elas se tornaram os principais alvos para o planejamento de fármacos associados a doenças do SNC.<sup>70</sup> Contudo, a não obtenção das estruturas cristalográficas de várias de suas proteínas incorre no desconhecimento do mecanismo das suas funções no organismo, e isso é um dos grandes gargalos que muitas vezes impedem o avanço das pesquisas focadas em receptores GPCRs.

O receptor GPCR 5-HT<sub>6</sub> faz parte do grupo de alvos biológicos, cujas estruturas tridimensionais ainda permanecem desconhecidas. No entanto, devido ao seu papel funcional perante a déficits cognitivos, o número de pesquisas envolvendo esse receptor tem aumentado enormemente. Para prever a interação do 5-HT<sub>6</sub> com seus ligantes, os pesquisadores lançam mão de abordagens computacionais<sup>74,102-105</sup> tendo como ponto de partida, a predição das suas coordenadas tridimensionais usando algum dos métodos supracitados.

Para prever estrutura tridimensional de uma determinada proteína, vários servidores são disponibilizados na WEB para esse propósito, dos quais podemos citar o RaptorX<sup>106</sup>, servidor que se distingue dos demais pela acurácia dos resultados obtidos. O RaptorX melhora a qualidade do alinhamento entre o alvo de interesse e os *templates* selecionados mesmo quando há baixa similaridade entre eles, além de avaliar o resultado do alinhamento.<sup>107</sup> Diante dos ótimos resultados, para se atingir os objetivos (Seção 3) desta pesquisa, esse servidor foi o escolhido para prever as coordenadas tridimensionais do receptor 5-HT<sub>6</sub> e assim explorar a interação entre ele e os ligantes selecionados (Seção 4).

## 2.2 Abordagem molecular focada no ligante

Na ausência de informações experimentais acerca do alvo biológico, a comunidade científica emprega a abordagem baseada no ligante (LBDD) para o planejamento de fármacos. Para esse fim, o método conhecido como Relação Quantitativa Estrutura-Atividade (QSAR) é o mais amplamente usado no campo de LBDD, uma vez que a partir da aplicação desse método é possível obter importantes informações das partes das moléculas que mais contribuem para o efeito biológico pretendido, reduzindo os custos e o tempo associados à realização de experimentos.

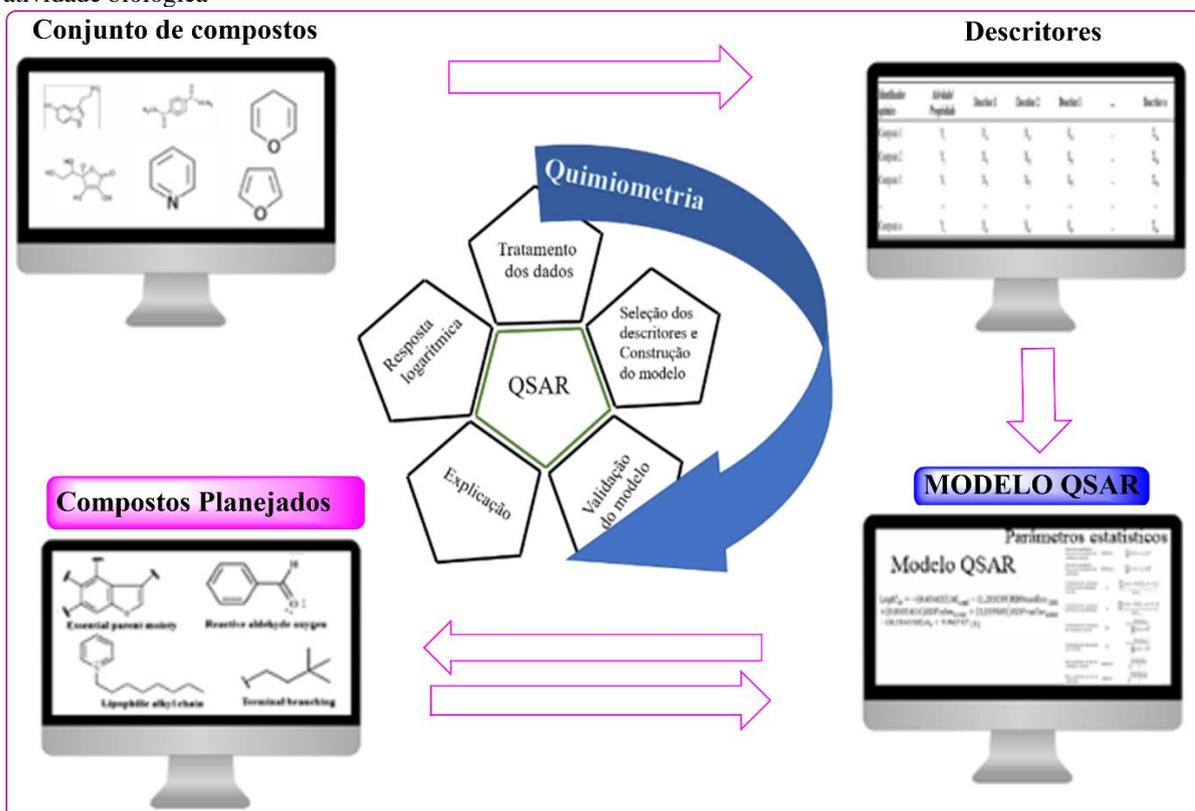
### 2.3 Relação Quantitativa Estrutura-Atividade-QSAR

Métodos de QSAR são ferramentas computacionais usadas há muito tempo, possuindo uma vasta aplicabilidade, por exemplo, na academia, na indústria, instituições governamentais, entre outras. No entanto, o seu maior emprego ocorre na área de química medicinal.<sup>108</sup>

Essas ferramentas são bastante úteis para entender e explicar o mecanismo químico molecular entre o fármaco e o seu alvo. Uma vez entendido o mecanismo que rege essa interação, a etapa seguinte é planejar e propor novos compostos que possam interagir eficazmente com o alvo molecular e assim resultar no efeito biológico esperado.<sup>108</sup>

De forma simplificada, o procedimento para gerar modelos em QSAR passa pelas seguintes etapas: **a)** os compostos são selecionados e otimizados para que descritores (moleculares, constitucionais, geométricos, eletrônicos, etc.) sejam obtidos; **b)** os descritores são selecionados para a construção do modelo; **c)** O modelo QSAR é construído e sua robustez é avaliada; **d)** a partir das análises do modelo obtido, novas moléculas são planejadas; **e)** o modelo QSAR gerado é utilizado para prever a atividade biológica dessas novas moléculas (Figura 19).

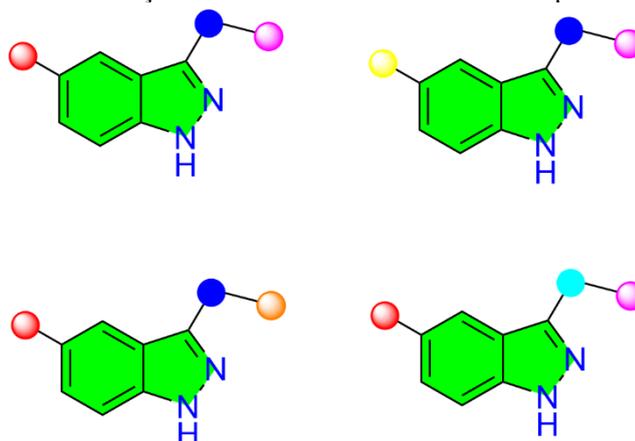
**Figura 19** - Etapas seguidas para a construção de modelos QSAR- seleção de um conjunto de compostos, cálculo e seleção de descritores, construção e validação do modelo QSAR, planejamento de compostos e predição da atividade biológica



Fonte: autoria própria

O objetivo da construção de modelos QSAR é relacionar a estrutura química à atividade biológica de grupos de compostos com *scaffold* comum e assim identificar qual parte da molécula mais contribui para desencadear o efeito biológico. Para isso, os substituintes são alterados ou removidos para verificar a importância deles (a Figura 20 apresenta um composto hipotético, no qual cada esfera colorida representa um substituinte diferente).

**Figura 20** - Modificações estruturais realizadas em um composto hipotético



Fonte: autoria própria

Matematicamente, a relação entre as propriedades estruturais (descritores) e a medida biológica ocorre aplicando-se a seguinte equação (Eq. 7):

$$\textit{Atividade Biológica} = f(\textit{propriedades estruturais}) \quad (7)$$

A estrutura química está relacionada à forma como os compostos são constituídos (tipos de átomos) e dispostos no espaço tridimensional, já a atividade biológica é uma medida realizada em laboratório decorrente das reações ou interações químicas entre um determinado composto e seu alvo biológico, sendo essas as características primordiais analisadas para determinar se uma molécula é bioativa.<sup>108,109</sup>

Modelos QSAR não devem ser construídos indiscriminadamente. De acordo com a OECD (do inglês, *Organisation for Economic Co-operation and Development*) vários modelos QSAR foram gerados e utilizados em avaliações de produtos químicos e em várias outras áreas, mas nem todos são modelos robustos, por isso que, em 2004, essa entidade elaborou um documento de boas práticas em QSAR, e em 2007 esse documento foi atualizado. Portanto, essas práticas devem ser consideradas, quando se deseja construir modelos confiáveis.<sup>110</sup>

As práticas em QSAR geraram modelos de extrema previsibilidade e de grande importância na área de desenvolvimento de fármacos com vários casos de sucesso.<sup>111-117</sup> Tem-se como exemplo, o *Donepezil*<sup>118</sup>, que, como mencionado (Figura **4d**), é um fármaco usado como inibidor da acetilcolinesterase, aprovado pela FDA em 1996. Portanto, pode-se inferir, com base nas aplicações bem-sucedidas, que essa técnica é bastante confiável e pode ser aplicada em várias áreas para solucionar vários problemas.

### 3 OBJETIVOS

Planejar novos antagonistas do receptor 5-HT<sub>6</sub>, potencial alvo para a doença de Alzheimer, empregando diferentes ferramentas computacionais.

#### 3.1 *Objetivos Específicos*

- Identificar e otimizar moléculas com atividade antagonista diante do receptor 5-HT<sub>6</sub>;
- Calcular e selecionar descritores que mais contribuem para o entendimento do efeito antagonista dos compostos;
- Gerar modelos QSAR;
- Construir as coordenadas tridimensionais do receptor 5-HT<sub>6</sub> e usar o método *Dinâmica* molecular para refinar o modelo gerado;
- Simular o ambiente fisiológico do receptor 5-HT<sub>6</sub> por meio do método *Dinâmica* molecular;
- Analisar as interações entre os compostos e o receptor 5-HT<sub>6</sub> por meio da técnica de *Docking* molecular;
- Planejar novos compostos a partir das análises de estrutura-atividade e *Docking* molecular;
- Usar os modelos QSARs, previamente construídos, para prever a atividade biológica dos compostos planejados.

## 4 METODOLOGIA

A predição das coordenadas tridimensionais do receptor 5-HT<sub>6</sub> foi realizada no servidor RaptorX<sup>106</sup> e, posteriormente, o modelo 3D foi refinando via técnica de *Dinâmica* molecular.<sup>92</sup> Os modelos QSARs 2D e 3D foram gerados usando diferentes metodologias, tais como, os métodos PLS (Mínimos quadrados parciais, do inglês, *Partial Least Squares*)<sup>119</sup>, CoMFA<sup>120</sup> (Análise Comparativa de Campos de Interação Molecular, do inglês, *Comparative Molecular Field Analysis*) e ANN<sup>121</sup> (Redes Neurais Artificiais, do inglês, *Artificial Neural Network*).

Dois conjuntos de compostos foram selecionados<sup>115,122</sup> para a construção dos modelos QSARs. Para o primeiro conjunto (Tabela 2), os métodos PLS e CoMFA foram empregados, enquanto para o segundo conjunto (Tabela 3), o método ANN foi aplicado. Concomitante à construção dos modelos QSARs, a técnica de *Dinâmica* molecular foi aplicada para simular o ambiente fisiológico do receptor 5-HT<sub>6</sub> e para isso um sistema membrana + receptor + ligante foi construído, conforme descrito na Seção 4.5. Posteriormente, os compostos mais ativos e menos ativos de ambos os conjuntos foram escolhidos para proceder as simulações de *Docking* molecular. Por último, por meio das análises estrutura-atividade e interações receptor-ligante, novos antagonistas do receptor 5-HT<sub>6</sub> foram planejados e em seguida submetidos aos modelos QSARs, previamente gerados, para que seus valores de atividade biológica fossem preditos. Essas diferentes metodologias aplicadas aos dois conjuntos de dados mencionados podem aumentar a probabilidade de encontrar potenciais antagonistas do receptor 5-HT<sub>6</sub> via estudos experimentais.

### 4.1 Construção das coordenadas tridimensionais do receptor 5-HT<sub>6</sub>

Para a construção do modelo 3D do receptor 5-HT<sub>6</sub>, algumas etapas foram necessárias, a saber:

➤ A sequência primária do receptor 5-HT<sub>6</sub> foi obtida a partir do servidor NCBI (*National Center for Biotechnology Information*)<sup>123</sup> e posteriormente submetida ao servidor RaptorX.<sup>106</sup> Este servidor faz uma busca em banco de dados por estruturas homólogas ou parte de estruturas com características similares ao alvo de interesse;

- O modelo gerado no RaptorX<sup>106</sup> foi submetido ao servidor *GalaxyRefine*<sup>124</sup> para ser refinado e para isso, o método de *Dinâmica Molecular* foi empregado;
- Para a validação externa do modelo, o servidor SAVES por meio do programa PROCHEK<sup>125</sup> foi utilizado.

#### 4.2 Construção dos modelos QSARs (2D e 3D)

Inicialmente, um modelo QSAR-2D foi construído com um conjunto de moléculas (conjunto 1) e diversos descritores foram selecionados, tais como, descritores moleculares, topológicos, eletrônicos, índices topológicos, entre outros. Para avaliar os efeitos estéricos e eletrônicos dos antagonistas (conjunto 1) do receptor 5-HT<sub>6</sub>, um modelo QSAR-3D foi gerado por meio do método CoMFA. Tanto para a construção do modelo QSAR-2D quanto para o 3D, o método de regressão linear PLS foi usado. Para avaliar o comportamento não linear dos compostos do conjunto 2, o método ANN foi empregado para construir um modelo QSAR.

A aplicação dessas diferentes metodologias auxiliou no planejamento de novos compostos e na predição de seus valores de atividade biológica, além de indicar o perfil ADME/Tox (Absorção, Distribuição, Metabolismo e Excreção/Toxicidade) de ambos, viabilizando futuros estudos experimentais para certificar seus valores de afinidade biológica bem como suas propriedades farmacocinéticas e farmacodinâmicas.

##### 4.2.1 PLS (*Partial Least Squares*)

O método PLS foi introduzido nos anos 80<sup>119</sup>, tratando-se de um método bastante versátil, tendo sua aplicabilidade em vários campos da ciência por ser empregado em análise de dados multivariados. O foco do método está em encontrar o subespaço linear das variáveis explicativas, isto é, a importância de cada uma, dada em linguagem matemática, para a atividade biológica.<sup>126</sup>

No método PLS, os descritores são representados pelas matrizes **X**, ao passo que a atividade biológica é descrita pela matriz **Y**.<sup>127</sup> No método PLS, os descritores são representados pela matriz **X**, ao passo que a atividade biológica é descrita pela matriz **Y**. Essas duas matrizes são correlacionadas da seguinte forma (Eq. 8):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + f \quad \mathbf{8}$$

Onde  $\beta$  é o coeficiente de correlação e  $f$  se refere ao erro, pois é improvável que  $y$  seja expresso exatamente em termos de  $\mathbf{X}$ .

Para cada conjunto de  $n$  amostras ( $i = 1, 2, \dots, n$ ), temos (Eq. **9**):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + f_i \quad \mathbf{9}$$

Colocando em termo de Matriz, temos (Eq. **10**):

$$y = X \beta + f \quad \mathbf{10}$$

A matriz  $X$  é apresentada da seguinte forma (Eq. **11**):

$$X = U R V^T \quad \mathbf{11}$$

L = V  
T= UR

Em que  $U$  é a matriz de score,  $V$  é a matriz de *Loading* e  $R$  é uma matriz bidiagonal (isto é, os elementos acima da diagonal principal são diferentes de zero).

Substituindo a equação **11** em **10** temos (Eq. **5**):

$$\beta = V R^{-1} U^T y \quad \mathbf{12}$$

Finalmente, é possível prever o valor de  $y$  para uma nova amostra a partir do seu valor de  $X$  (Eq. **13**):

$$y_{predito} = X_{novo} \beta = X_{novo} V R^{-1} U^T y \quad \mathbf{13}$$

O PLS encontra o número ótimo de variáveis latentes (**VL**) por meio de um procedimento denominado de validação cruzada (*Cross-Validation*). A validação cruzada avalia a capacidade de generalização do modelo de um conjunto de dados. Nesse procedimento, o conjunto de dado

é dividido em subgrupos de tamanho  $N$  para a construção de vários modelos. O cálculo é executado sempre com a exclusão de um dos subgrupos para impedir o super ajuste (*Overfitting*). Todo esse procedimento é realizado para encontrar modelos que sejam acurados e confiáveis e que tenham eficiência diante da inclusão de novas amostras.<sup>128</sup> A Tabela 1 mostra os parâmetros estatísticos avaliados que direciona a escolha do número ótimo de variáveis latentes, e, conseqüentemente, obtenção de modelos confiáveis.

**Tabela 1** - Parâmetros estatísticos analisados para validar internamente o modelo QSAR-2D

Parâmetros estatísticos	Descrição
$Press = \sum (y_{iexp} - \hat{y}_{ival})^2$	Soma dos quadrados dos erros de previsão.
$SEV = \left[ \frac{PRESS}{I} \right]^{\frac{1}{2}}$	Erro padrão de validação.
$SEC = \left[ \frac{\sum (y_{ical} - \hat{y}_{ical})^2}{I - VL - 1} \right]^{\frac{1}{2}}$	Erro padrão de calibração.
$q^2 = \frac{\sum (y_{iexp} - \hat{y}_{ival})^2}{\sum (y_{iexp} - y_{mean})^2}$	Coefficiente de correlação de validação cruzada.
$r^2 = \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - \hat{y}_{mean})^2}$	Coefficiente de determinação múltipla.

$y_{iexp}$  representa o valor experimental de Y da  $i$ -ésima amostra;

$\hat{y}_{ival}$  representa o valor predito de Y pelo modelo construído sem a amostra  $i$ ;

$I$  é o número de amostras do conjunto utilizado na validação cruzada;

$\hat{y}_{ical}$  representa o valor calculado de Y da  $i$ -ésima amostra com todas as amostras do conjunto;

$y_{mean}$  representa o valor médio de Y.

#### 4.2.2 Análises Comparativa de Campos Moleculares (CoMFA)

É importante construir modelos QSAR-3D, uma vez que importantes informações (como propriedades estéricas e eletrostáticas) são extraídas a partir das análises entre a estrutura tridimensional e o efeito biológico intrínseco dos compostos selecionados para o estudo. Um dos métodos QSAR-3D amplamente empregado é o CoMFA, desenvolvido por Cramer e colaboradores.<sup>120</sup> Nesse método, os campos moleculares (campos estéricos e eletrostáticos) são analisados e relacionados à resposta biológica, partindo-se de uma premissa fundamental - em nível molecular, as interações não-covalentes são as responsáveis por produzir o efeito biológico requerido. Para se construir modelos QSAR-3D usando o método CoMFA, todos os ligantes de interesse são colocados em uma caixa retangular e por meio de uma sonda

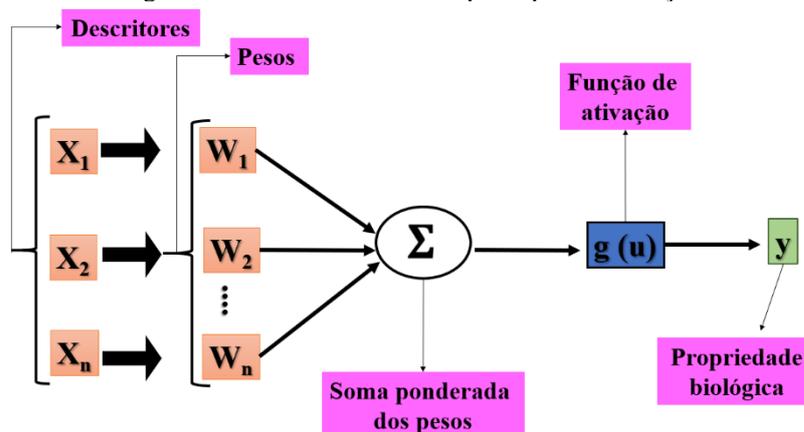
introduzida em cada ponto de intersecção da caixa, a energia de interação entre os ligantes e a sonda é calculada em cada ponto. O resultado final corresponde às propriedades eletrostáticas (Coulombic) e estéricas (van der Waals), as quais serão tidas como os descritores para a construção do modelo QSAR-3D. Estes descritores são correlacionados à atividade biológica por meio do método PLS (também empregado para o desenvolvimento de modelos QSAR-2D), o qual indica quais as contribuições eletrostáticas e estéricas são favoráveis ou desfavoráveis para a propriedade biológica de interesse.<sup>120</sup>

#### 4.2.3 *Redes Neurais Artificiais*

Aprendizagem de máquina (do inglês, *Machine Learning*) é uma das abordagens computacionais mais céleres para se obter os resultados pretendidos, além de ser eficiente e apresentar baixo custo computacional. A relação entre a estrutura química e a atividade biológica dos compostos são processadas via algoritmo de reconhecimento padrão, e o modelo obtido é posteriormente utilizado para prever a resposta biológica de novas entidades químicas. Convencionalmente, essa relação era avaliada por meio das análises de Hansch, os quais são modelos de regressão linear. No entanto, a relação entre essas variáveis nem sempre são tidas como lineares, o que dificulta uma análise mais assertiva. Sendo assim, Redes Neurais Artificiais é um método amplamente empregado para modelar variáveis, cuja relação pode não ser linear.<sup>129</sup>

ANNs são modelos matemáticos inspirados no funcionamento do neurônio biológico, em que os sinais de propagação são ativados via neurotransmissão (interação entre as células neurais).<sup>130</sup> Basicamente, cada neurônio recebe vários sinais de entrada (para o nosso problema, esses sinais são os descritores) e a esses descritores são atribuídos pesos, os quais são processados por meio de uma função ponderada (função peso). O produto das entradas pela soma ponderada dos pesos é submetido a uma função de transferência, também conhecida como função de ativação.<sup>131</sup> Estas são as responsáveis por permitir a obtenção da resposta (para nosso problema, predição da atividade biológica). Caso a resposta seja ruim, isto é, caso a rede não tenha sido bem treinada, todo o processo é refeito, os pesos são atualizados até que se obtenham uma boa resposta, e assim um modelo robusto e com poder de previsibilidade satisfatório pode ser encontrado (Figura 21).

**Figura 21** - Estrutura geral de uma Rede Neural adaptada para construção de modelos QSAR



Fonte: autoria própria

ANNs aprende a relação entre *input* (descritores) e *output* (propriedade biológica,  $pK_i$ ) por experiência, cujo treinamento ocorre por meio de um determinado algoritmo de aprendizagem. Resumidamente, esse treinamento ocorre da seguinte forma: a atividade neuronal é estabelecida por pesos sinápticos que determina a resposta biológica ( $pK_i$ ) e então, a partir daí, o erro é conhecido. Se os erros alcançados forem demasiados, podemos tentar reduzi-los fazendo pequenas alterações nos pesos sinápticos (Eq.14) e posteriormente, a rede é retreinada para se obter um modelo robusto.<sup>132,133</sup>

$$\Delta W_{i,j,n} = F_n + \alpha W_{i,j,n-1} \quad 14$$

Onde  $\Delta W_{i,j}$  se trata da modificação no peso para cada nó da rede,  $\alpha$  é o fator momentum e  $F$  é a função peso atualizada, cujo papel é informar como os pesos são alterados ao longo do treinamento.<sup>134</sup> Para essa pesquisa, o algoritmo de Levenberg-Marquardt (L-M) foi utilizado como função de atualização de peso, o qual interrompe automaticamente o treinamento da rede quando há aumento exacerbado no erro.<sup>135</sup>

### 4.3 Conjunto de dados

Dois conjuntos de compostos<sup>115,122</sup> foram escolhidos para a construção dos modelos QSAR. Para o primeiro conjunto (Tabela 2), os modelos QSAR 2D e 3D foram construídos usando os métodos PLS e CoMFA.

**Tabela 2 - Conjunto 1:** estrutura e atividade biológica dos antagonistas do receptor 5-HT<sub>6</sub>

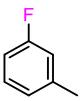
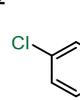
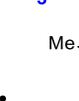
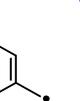
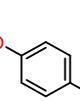
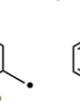
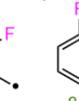
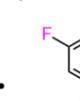
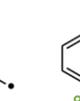
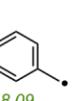
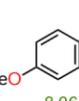
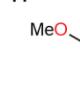
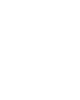
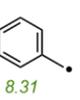
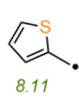
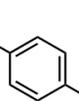
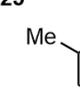
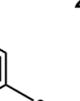
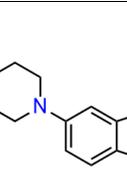
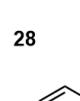
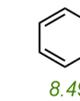
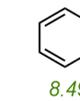
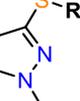
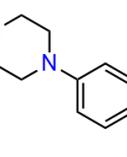
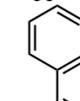
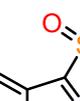
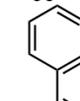
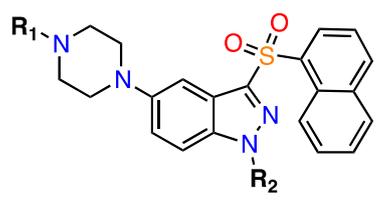
 <p>Intervalo de valores de <math>pK_i</math>: 6.87–8.87</p>	<p><b>1</b></p>  <p>6.87</p>	<p><b>2</b></p>  <p>6.90</p>	<p><b>3</b></p>  <p>8.87</p>	<p><b>4</b></p>  <p>7.06</p>	<p><b>5</b></p>  <p>7.32</p>
 <p>Intervalo de valores de <math>pK_i</math>: 7.92–9.05</p>	<p><b>6</b></p>  <p>8.72</p>	<p><b>7</b></p>  <p>8.26</p>	<p><b>8</b></p>  <p>8.77</p>	<p><b>9</b></p>  <p>8.28</p>	<p><b>10</b></p>  <p>8.40</p>
	<p><b>11</b></p>  <p>9.00</p>	<p><b>12</b></p>  <p>8.09</p>	<p><b>13</b></p>  <p>8.96</p>	<p><b>14</b></p>  <p>7.92</p>	<p><b>15</b></p>  <p>9.05</p>
	<p><b>16</b></p>  <p>8.31</p>	<p><b>17</b></p>  <p>8.11</p>	<p><b>18</b></p>  <p>8.59</p>	<p><b>19</b></p>  <p>8.80</p>	<p><b>20</b></p>  <p>8.96</p>
 <p>Intervalo de valores de <math>pK_i</math>: 6.96–8.34</p>	<p><b>21</b></p>  <p>7.96</p>	<p><b>22</b></p>  <p>6.96</p>	<p><b>23</b></p>  <p>7.57</p>	<p><b>24</b></p>  <p>6.91</p>	<p><b>25</b></p>  <p>6.68</p>
 <p>Intervalo de valores de <math>pK_i</math>: 8.49–8.80</p>	<p><b>26</b></p>  <p>7.66</p>	<p><b>27</b></p>  <p>8.34</p>	<p><b>28</b></p>  <p>8.49</p>	<p><b>29</b></p>  <p>8.54</p>	<p><b>30</b></p>  <p>8.77</p>
 <p>Intervalo de valores de <math>pK_i</math>: 8.54–8.72</p>	<p><b>31</b></p>  <p>8.80</p>	<p><b>32</b></p>  <p>8.77</p>	<p><b>33</b></p>  <p>8.72</p>		

Tabela 2- Continuação

	R <sub>2</sub>	R <sub>1</sub> = Me	R <sub>1</sub> = H
		8.74 ( <b>34</b> )	8.54 ( <b>35</b> )
		8.74 ( <b>36</b> )	8.54 ( <b>37</b> )
		8.85 ( <b>38</b> )	8.34 ( <b>39</b> )
	8.70 ( <b>40</b> )	8.12 ( <b>41</b> )	

				
7.80	7.66	8.30	7.96	7.92
				
8.05	7.47	8.00	8.10	8.52
				
9.00	8.70	7.46	8.15	7.62

		
7.44	7.24	6.88

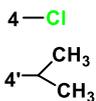
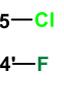
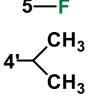
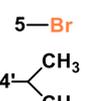
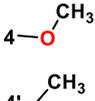
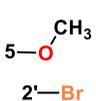
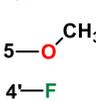
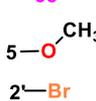
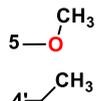
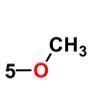
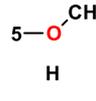
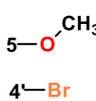
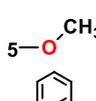
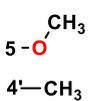
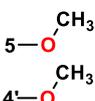
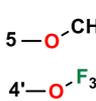
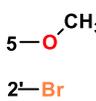
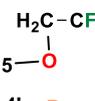
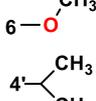
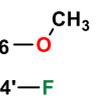
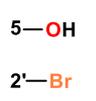
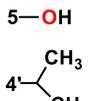
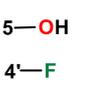
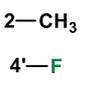
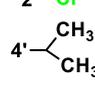
\*Os compostos do conjunto teste estão destacados em azul.

Fonte: Da Silva e colaboradores.<sup>136</sup>

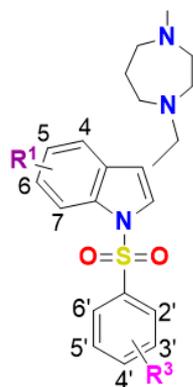
A Tabela 3 apresenta o conjunto de compostos selecionados para a construção do modelo ANN. Para garantir que o modelo seja estatisticamente robusto, esses compostos foram divididos em conjunto treino, teste e validação. Em adição, para comprovar o poder preditivo do modelo, além do conjunto teste, um conjunto externo (Tabela 4) foi selecionado a partir do banco de dados ChEMBL<sup>100</sup> e o modelo ANN foi usado para prever os valores de atividade biológicas desses compostos.



Tabela 3- Continuação

<p>24</p>  <p>8.04</p>	<p>25</p>  <p>7.51</p>	<p>26</p>  <p>7.63</p>	<p>27</p>  <p>7.76</p>	<p>28</p>  <p>7.45</p>	<p>29</p>  <p>7.22</p>	<p>30</p>  <p>8.40</p>	<p>31</p>  <p>7.92</p>	<p>32</p>  <p>8.04</p>	<p>33</p>  <p>7.77</p>
<p>34</p>  <p>7.56</p>	<p>35</p>  <p>7.94</p>	<p>36</p>  <p>7.84</p>	<p>37</p>  <p>7.74</p>	<p>38</p>  <p>8.70</p>	<p>39</p>  <p>8.79</p>	<p>40</p>  <p>8.30</p>	<p>41</p>  <p>8.37</p>		
<p>42</p>  <p>7.99</p>	<p>43</p>  <p>8.11</p>	<p>44</p>  <p>7.81</p>	<p>45</p>  <p>7.74</p>	<p>46</p>  <p>7.42</p>	<p>47</p>  <p>7.44</p>	<p>48</p>  <p>7.02</p>	<p>49</p>  <p>7.44</p>		
<p>50</p>  <p>6.62</p>	<p>51</p>  <p>7.73</p>	<p>52</p>  <p>7.30</p>	<p>53</p>  <p>7.46</p>	<p>54</p>  <p>6.70</p>	<p>55</p>  <p>6.71</p>	<p>56</p>  <p>7.25</p>	<p>57</p>  <p>7.53</p>		

Compostos 58 a 60



Intervalo de valores de  $pK_a=8.11$  a  $8.70$

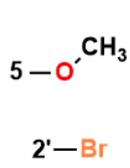
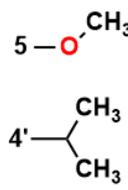
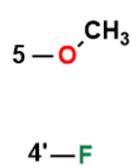
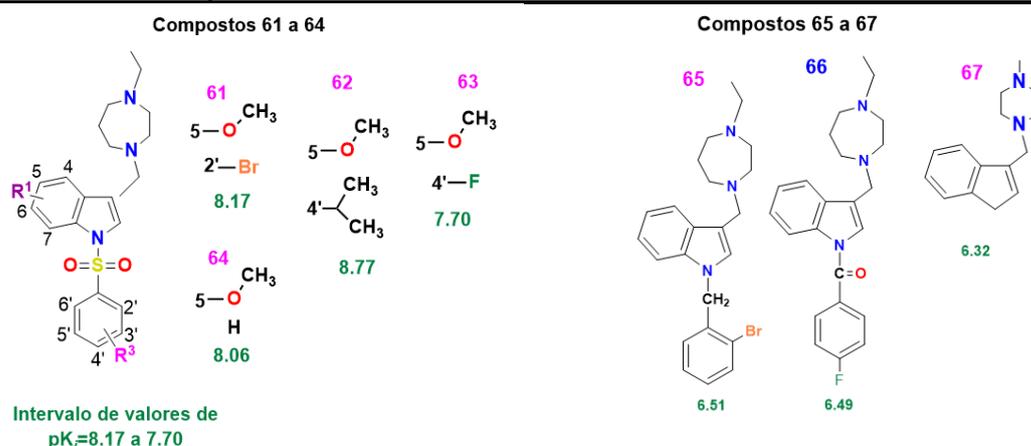
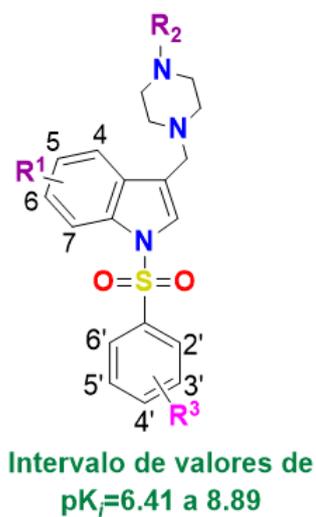
<p>58</p>  <p>8.24</p>	<p>59</p>  <p>8.70</p>	<p>60</p>  <p>8.11</p>
---	--	---

Tabela 3- Continuação



Compostos 68 a 76



	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	pK <sub>a</sub>
68	H	H	H	8.12
69	H	2'-Br	H	8.35
70	5-O-CH <sub>3</sub>	H	2'-Br	8.89
71	5-O-CH <sub>3</sub>	H	4'-CH(CH <sub>3</sub> ) <sub>2</sub>	8.24
72	5-O-CH <sub>3</sub>	H	4'-F	8.16
73	H		2'-Br	6.44
74	H		2'-Br	6.41
75	H		2'-Br	6.88
76	5-O-CH <sub>3</sub>	H <sub>2</sub> C-F	-CH <sub>2</sub>	6.46

Tabela 3- Continuação

Compostos 77 e 78		Compostos 79 a 81						
Intervalo de valores de $pK_f=6.40$ a $6.54$		Intervalo de valores de $pK_f=6.46$ a $7.02$						
<u>77</u>	<u>78</u>	$R_1$	$R_2$	$R_3$	X	Y	Z	$pK_f$
5-Br	5-Br	H	H	2'-Br	N	CH	CH	7.02
2'-Br	3'-Br	H	CH <sub>3</sub>	H	CH	N	CH	6.58
6.40	6.54	5-O-CH <sub>3</sub>	CH <sub>3</sub>	2'-Br	CH	CH	N	6.46

Fonte: Da Silva e colaboradores.<sup>137</sup>

Para certificar o poder preditivo do modelo ANN-QSAR, os seguintes compostos foram selecionados (Tabela 4):

Tabela 4 - Estrutura molecular dos compostos usados para verificar o poder preditivo do modelo ANN-QSAR

ID ChEMBL		$R_1$	$R_2$	$pK_f$
2159470				6.59
2159464				6.06
2159456				5.56
60669				8.00
434365				7.42
378656				7.77

Tabela 4- Continuação

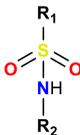
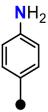
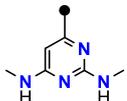
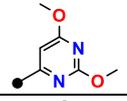
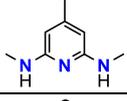
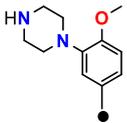
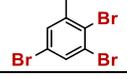
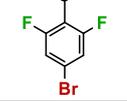
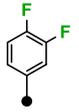
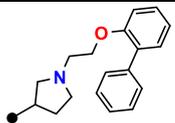
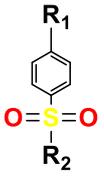
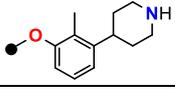
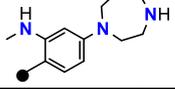
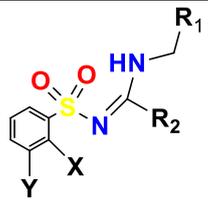
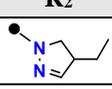
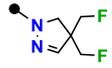
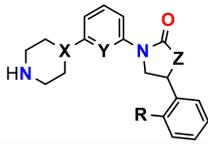
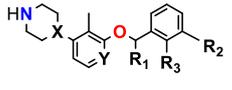
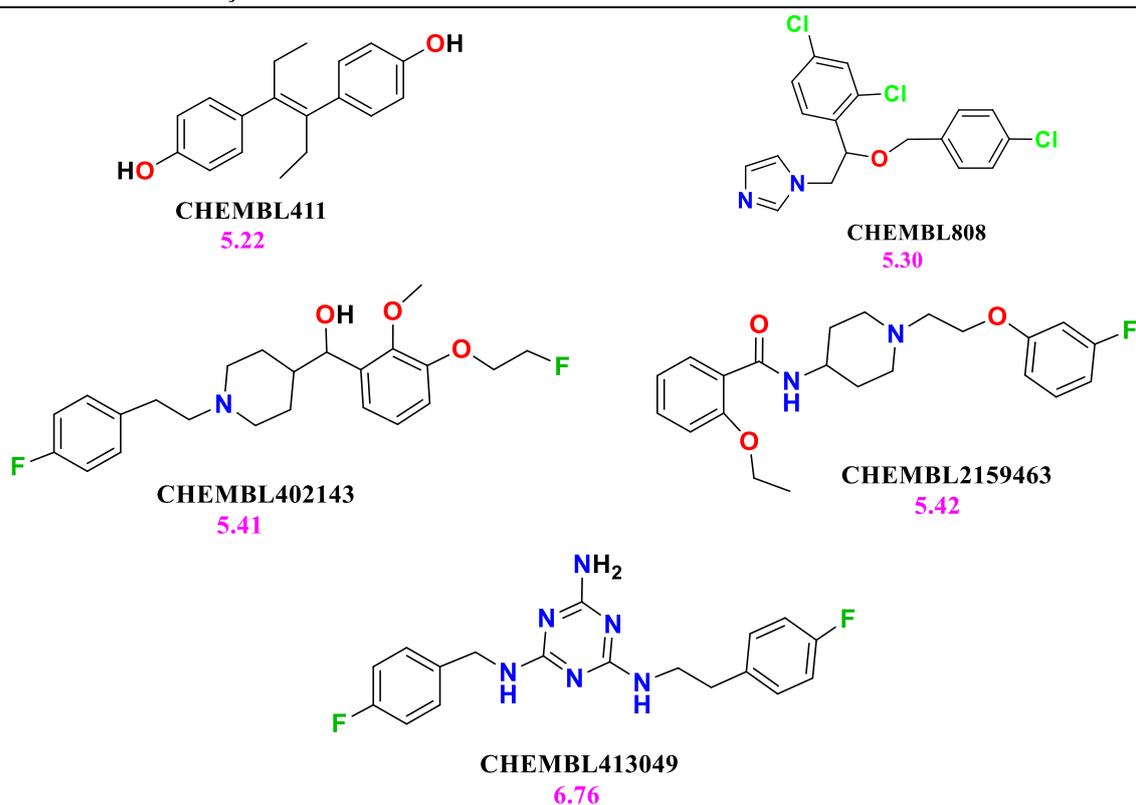
ID ChEMBL		R <sub>1</sub>	R <sub>2</sub>			pK <sub>i</sub>	
433461						7.33	
62193						6.21	
46071						7.92	
154641						8.60	
347335						6.80	
2159481						6.49	
24474		NH <sub>2</sub>				9.94	
482748		CH <sub>3</sub>				7.54	
95136		F				8.85	
ID ChEMBL		R <sub>1</sub>	R <sub>2</sub>	X	Y	pK <sub>i</sub>	
1834222				Cl	H	7.04	
1834351		CH <sub>3</sub>		H	Cl	7.04	
ID ChEMBL		X	Y	Z	R	pK <sub>i</sub>	
3260792		N	N	CH <sub>2</sub>	H	7.32	
3260806		N	CH	O	Cl	5.72	
3260790		CH	CH	CH <sub>2</sub>	H	7.16	
ID ChEMBL		R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	X	Y	pK <sub>i</sub>
484164		CH <sub>3</sub>	Cl	H	CH	CH	8.55
506736		CH <sub>3</sub>	F	H	CH	CH	8.16
482990			F	H	CH	N	8.37
483559		H	H	F	N	CH	7.94

Tabela 4 – Continuação



Fonte: Da Silva e colaboradores.<sup>137</sup>

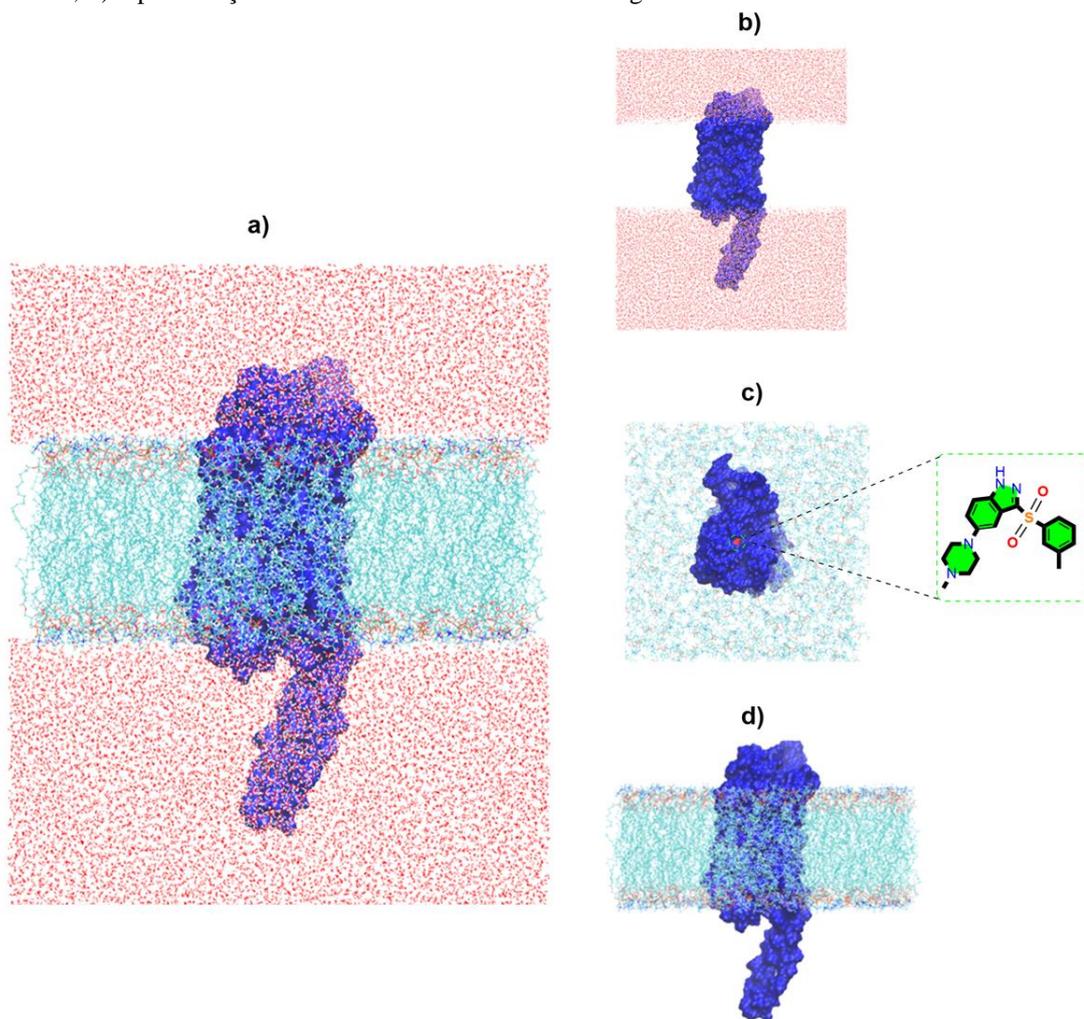
#### 4.4 Simulação de Docking molecular

As simulações de *Docking* molecular foram realizadas no Glide<sup>138</sup> (pacote incluído no programa Maestro) com o sistema receptor e os ligantes mais ativos e menos ativos, respectivamente, de ambos os conjuntos (C15 e C25- conjunto 1, C70 e C67- conjunto 2). O módulo *Protein Preparation Wizard* (PPW) foi usado para adicionar hidrogênios, manter o pH em 7 e adicionar cargas aos resíduos. O sítio ativo do receptor 5-HT<sub>6</sub> foi definido usando o módulo *Receptor Grid Generation* e uma caixa de tamanho 10 x 10 x 10 foi construída. Para essa simulação, a função de score *XP extra precisão* foi escolhida e ao final, uma avaliação dos principais tipos de interações entre o receptor e os compostos mencionados foi realizada.

#### 4.5 Simulações de Dinâmica molecular

O receptor 5-HT<sub>6</sub> ancorado ao ligante C15 do conjunto 1 foi embebido dentro de uma membrana para simular o seu ambiente fisiológico (Figura 22). Posteriormente, o sistema foi submetido ao procedimento de *Dinâmica* molecular.

**Figura 22** - Simulação de *Dinâmica* molecular- **a)** Representação de todo o sistema: membrana, moléculas de água, receptor e ligante; **b)** representação do sistema sem a membrana POPC; **c)** visualização do receptor, ligante e membrana; **d)** representação do sistema sem as moléculas de água



**Fonte:** simulação realizada em colaboração com o profº Guedmiller Souza de Oliveira da Universidade Federal de Uberlândia.

A membrana selecionada para simular o ambiente biológica foi a do tipo POPC (fosfatidilcolina, nome completo- 1-plalmiloil-2-oleoil-sn-glicerol-3-fosfocolina) 90° x 90° e o modelo de solvatação usado na simulação foi do tipo TIP3P (do inglês, *Transferable Intermolecular Potentials*). Para a simulação, o programa NAMD versão **2.14b1**<sup>139</sup> foi usado, e as seguintes condições periódicas de contorno foram empregadas:(i) Pistão Langevin<sup>140</sup> para

ajuste de pressão em 1,0 atm; **(ii)** *ensemble* do tipo NVT a 290 K (termostato Langevin) durante 5 ns; **(iii)** para correções de longo alcance, um corte de distância de 12 Å usando o formalismo de soma de Ewald<sup>141</sup> foi utilizado. A estabilidade numérica durante a trajetória foi alcançada utilizando o ponto de corte mencionado. Após a primeira etapa **(i)**, um equilíbrio NPT foi executado ao longo de 20 ns. **(iv)** O volume médio obtido em **(iii)** foi usado para realizar a simulação usando o *ensemble* NVT para obter uma nova trajetória em 5 ns. Todos os componentes foram parametrizados empregando-se o protocolo de campo de força CHARMM36.<sup>142</sup> Em seguida, os resultados foram analisados usando o software VMD.<sup>143</sup>

#### 4.6 Procedimentos

Para alcançar os objetivos pretendidos, seguimos os seguintes procedimentos computacionais:

- As estruturas dos compostos (Tabelas 2 e 3) foram construídas e otimizadas por meio do pacote Gaussian 09<sup>144</sup>, utilizando o método DFT (do inglês, *Density Functional Theory*). O B3LYP<sup>145</sup> e 6-31G (d,p)<sup>146</sup> foram o funcional e as funções de base, respectivamente, escolhidas para proceder com os cálculos das propriedades eletrônicas das estruturas dos compostos;
- Os descritores eletrônicos, como energia dos orbitais HOMO (Orbital de maior energia ocupado, do inglês - *Highest Occupied Molecular Orbital*) e do LUMO (Orbital de menor energia desocupado, do inglês - *Lowest Unoccupied Molecular Orbital*), foram extraídos a partir dos resultados dos cálculos realizados na etapa anterior;
- Os descritores topológicos, geométricos e constitucionais foram calculados por meio do *software* e-Dragon<sup>147</sup>;
- Os descritores foram previamente selecionados por meio da aplicação de uma matriz de correlação (correlação entre os descritores e a propriedade biológica). Aqueles que apresentaram coeficiente de correlação de Pearson menor que 0,3 (em módulo) foram excluídos;

- A técnica de algoritmo genético<sup>148</sup> foi utilizada para realizar uma segunda seleção de descritores. A partir daí, selecionamos os descritores que mais contribuem para a propriedade biológica, mas na etapa que antecede a construção do modelo final, uma análise de autocorrelação entre eles foi realizada para excluir um dos descritores que contém a mesma informação, e assim evitar redundâncias;
  
- Os modelos QSARs foram construídos por meio dos *softwares* Pirouette versão **3.10**<sup>127</sup>, Sybyl **18.2**<sup>149</sup> e MatLab **R2015a**<sup>150</sup>;
  
- Todos os modelos foram avaliados interno e externamente para garantir a robustez e o poder preditivo de ambos;
  
- O modelo tridimensional do receptor 5-HT<sub>6</sub> foi construído no servidor RaptorX<sup>106</sup> e posteriormente, refinado por meio da técnica de *Dinâmica* molecular;
  
- Um sistema com o receptor, ligante, moléculas de água, íons e a membrana biológica foi construído para simular o ambiente fisiológico do receptor;
  
- A técnica de *Docking* molecular foi aplicada por meio do programa Glide<sup>138</sup> para avaliar as principais interações entre o receptor e os ligantes, estes foram escolhidos com base nos seus valores de pK<sub>i</sub>;
  
- A partir das análises dos modelos QSARs e das informações extraídas das interações receptor-ligante, novos antagonistas do receptor 5-HT<sub>6</sub> foram planejados;
  
- Os novos compostos foram submetidos a análises do perfil ADME/Tox para avaliar suas propriedades *drug-likeness* por meio do módulo QikProp<sup>151</sup> do pacote Schrödinger. A acessibilidade sintética desses compostos também foi explorada pelo servidor SwissADME<sup>152</sup> a fim de avaliar o grau de dificuldade para sintetizá-los.

## 5 RESULTADOS E DISCUSSÕES

### 5.1 Construção das coordenadas tridimensionais do receptor 5-HT<sub>6</sub>

O servidor RaptorX<sup>106</sup> para construir um modelo tridimensional de um determinado alvo biológico faz uma varredura nos principais bancos de dados para encontrar homólogos da sequência *query* (alvo a ser modelado). A partir daí, um alinhamento é realizado para identificar as partes mais conservadas do alvo de interesse quando comparado a outros alvos de uma mesma família (para o estudo em questão o alvo é 5-HT<sub>6</sub>, família GPCR). No entanto, vale salientar que o servidor RaptorX<sup>106</sup> usa diferentes metodologias, como por exemplo aprendizagem de máquina para fazer essa varredura, e, portanto, a busca por alvos homólogos para realizar o alinhamento não se limita a alvos de uma mesma família. A Figura 23 apresenta um tipo de alinhamento entre o receptor 5-HT<sub>6</sub> e alguns outros alvos pertencentes a família GPCR, em que nas regiões conservadas o grau de similaridade é maior.

**Figura 23** - Alinhamento realizado entre a sequência primária do receptor 5-HT<sub>6</sub> e alguns outros alvos GPCRs. Nesse alinhamento, a região em vermelho são os resíduos conservados, aqueles que estão na mesma posição entre todos os alvos biológicos

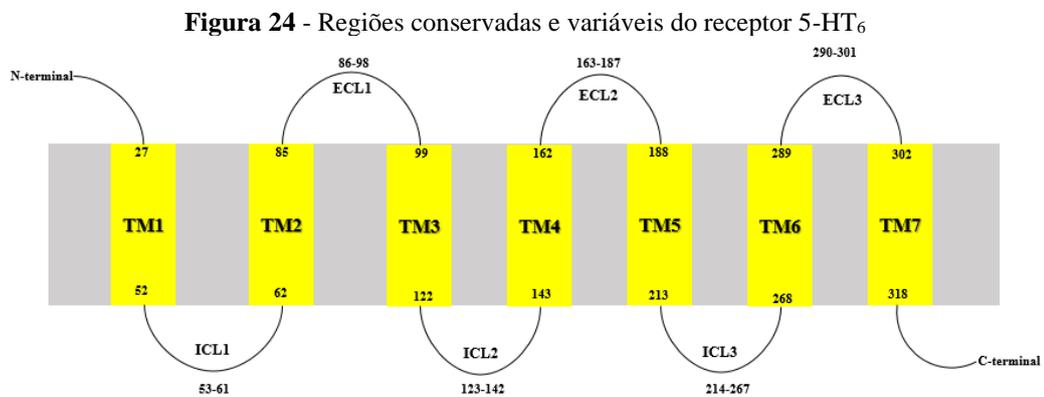
```

3kj6A -----GMGIVMSL-IVLAIVFGNVLVITAIA-FERLQTVNRYFITSL
2r4rA -----GIVMSL-IVLAIVFGNVLVITAIA-FERLQTVNRYFITSL
2r4sA -----GIVMSL-IVLAIVFGNVLVITAIA-FERLQTVNRYFITSL
41b4A -----EEQGNKLHWAALLILMVIIPTIGGNTLVILAVSLEKKLQYATNYFLMSL
5HT6  MVPEPGPTANSTPANGAGPPSAPGGSGWVAALCVVIALTAAANSLIILALICTQPALRNTSNFFLVSL
3kj6A  ACADLVMSGLAVVPPFGA-----FWTSIDVLCVTASIEETLCVIAVDRYFAITSPFKYQ
2r4rA  ACADLVMSGLAVVPPFG-----FWTSIDVLCVTASIEETLCVIAVDRYFAITSPFKYQ
2r4sA  ACADLVMSGLAVVPPFG-----FWTSIDVLCVTASIEETLCVIAVDRYFAITSPFKYQ
41b4A  AVADLLVGLFVMPIALLTIMFEAMWPLPLVLCPAWLFDLVLFSTASINWHLCAISVDRYIAIKKPIQAN
5HT6  FTSDLMSGLVMPFAMLNALY-GRWVLRGLCLLWTFDVMCCSASILNLCILSDRYLLLSPLRYK
3kj6A  SLLTKNKARVIIILMVWIVSGLT-----SSIVSFYVPLVI
2r4rA  SLLTKNKARVIIILMVWIVSGLT-----SSIVSFYVPLVI
2r4sA  SLLTKNKARVIIILMVWIVSGLT-----SSIVSFYVPLVI
41b4A  QNSRATAFIKITVWVWLSIGIAIPVPIKGIETN---PNNITCVLTKERFGDFMLFGSLAAFFTPLAI
5HT6  LRMTPLRALALVLGASLAALASFLPLLLGWHELGHARPPVPGQCRLLASLPFVLVASGLTFEFLPSGA
3kj6A  MVFVYSRVFQEAQRQLQKI-----
2r4rA  MVFVYSRVFQEAQRQLQKI-----
2r4sA  MVFVYSRVFQEAQRQLQKI-----
41b4A  MIVTYFLTIIHALQKKAADLEDNWETLNDNLKVEKADNAAQVKDALTKMRAAALDAQKDFRHGFDIL
5HT6  ICFTYCRILLAARKQAVQVASLTTGMSAQASETLQVPRTPRPGVESADSRRLATKHSRKALKASLTLG
3kj6A  -----DKSEGRFH-----VKFCLKEHKALKTLGIIIMGTFT
2r4rA  -----DKSEGRFH-----V-FCLKEHKALKTLGIIIMGTFT
2r4sA  -----DKSEGRFH-----V-FCLKEHKALKTLGIIIMGTFT
41b4A  VGQIDDKALKLANEGKVKEAQAAAQKLTTRNAYIQKYL-----QTISNEQRASKVLGIVFFFL
5HT6  ILLGMFFVTWLPFFVANIVQAVDCISPGFLFDVLTWLGVCNSTMNPPIIYPLFMRDFKRALGRFLPCPR
3kj6A  LCWLPFFI-----VYILLNWIGYVN-SGFNPLIYCRSPDFRIAFQELL-----
2r4rA  LCWLPFFI-----LNWIGYVN-SGFNPLIYCRSPDFRIAFQELL-----
2r4sA  LCWLPFFI-----LNWIGYVN-SGFNPLIYCRSPDFRIAFQELL-----
41b4A  LMCPPFFITNITLVLCDSNQTTLQMLLEIFVWIGYVSSGVNPLVYTLFNKTRFADFGRYI-----
5HT6  CPRERQASLASPSLRTSHSGPRPGLSLQOVLPLPLPPDSDSDSADSGSGSSGLRLTAQLLLPGEATQD
3kj6A  -----CLR--R-----SS--
2r4rA  -----CLR--R-----SSLK
2r4sA  -----CLR--R-----SSLK
41b4A  -----TCNYR-----
5HT6  PPLPTRA AAAAVNFFNIDPAEPELRPHPLGIPTN
    
```

Fonte: autoria própria

A Figura 23 apresenta o alinhamento entre alguns alvos, aqueles que apresentam grau de identidade maior, no entanto, vários outros alvos também foram usados para o alinhamento, e assim, informações confiáveis foram extraídas para a construção do modelo.

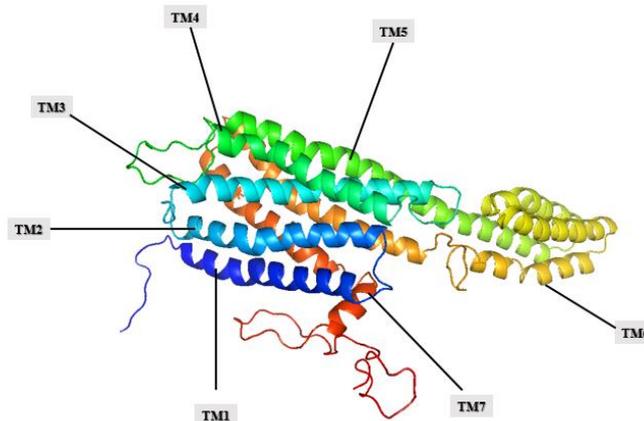
Para identificar as regiões conservadas, o método de vetores de suporte de máquina (SVM), técnica de aprendizagem de máquina, foi utilizado por meio do servidor MENSAT-SVM.<sup>153</sup> Essa técnica é utilizada para construir modelos de classificação, para esse caso, o modelo foi construído com várias proteínas pertencentes a uma mesma família, aumentando, dessa forma, o grau de confiabilidade nas informações extraídas. A Figura 24 mostra as regiões conservadas da proteína 5-HT<sub>6</sub> e o número de resíduos para cada região, além das regiões variáveis, *loop* intra (ICL) e extra (ECL) celular.



Fonte: autoria própria

Após o alinhamento, as coordenadas tridimensionais do receptor 5-HT<sub>6</sub> (Figura 25) foram construídas, mas uma parte das coordenadas não apresentou alinhamento satisfatório com outros alvos e por isso, optou-se por deixá-la de fora, sem prejuízo para o entendimento da interação receptor-ligante, uma vez que a parte excluída não faz parte do sítio de ligação.

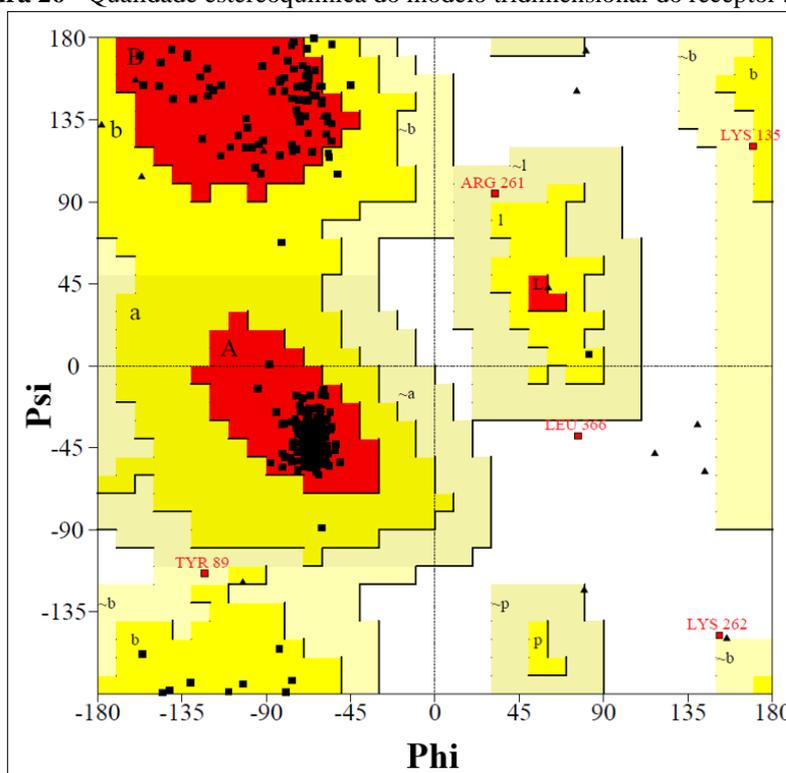
**Figura 25 - Modelo construído e suas regiões conservadas**



Fonte: autoria própria

O servidor SAVES foi utilizado para avaliar a qualidade do modelo obtido, esta ocorre por meio da análise do gráfico bidimensional Ramachandran (Figura 26). O diagrama Ramachandran mostra as regiões permitidas (regiões coloridas) e proibidas (regiões em branco) para todos os resíduos de aminoácidos. A análise da rotação desses ângulos de torções ( $\psi$  e  $\phi$ ) determina as regiões que são permitidas, em que o ângulo entre a ligação do carbono  $\alpha$  e o nitrogênio é referente ao ângulo *phi* ( $\phi$ ), ao passo que o ângulo *psi* ( $\psi$ ) é aquele entre a o carbono  $\alpha$  e o carbono da carbonila.

**Figura 26** - Qualidade estereoquímica do modelo tridimensional do receptor 5-HT<sub>6</sub>



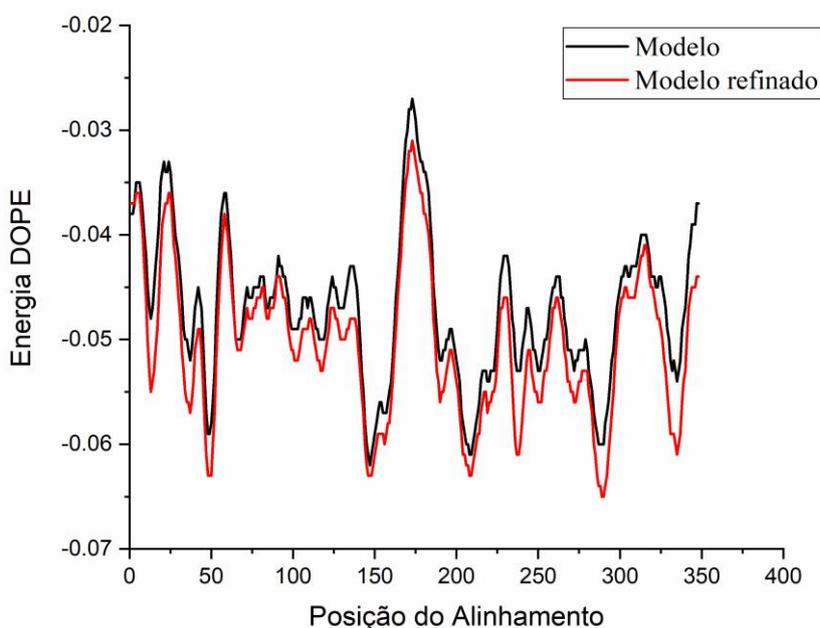
Fonte: Autoria própria

A análise da qualidade estereoquímica do modelo mostra que 99,4% dos resíduos de aminoácidos se encontram em regiões favorecidas, sendo que 93,7% deles estão em regiões muito favorecidas, 5,7% em regiões permitidas e 0,5% em regiões desfavorecidas (regiões em branco). Para que um modelo seja considerado confiável, é necessário que pelo menos 90% dos resíduos estejam em regiões favorecidas. Com base nessa recomendação, infere-se que o modelo gerado possui boa qualidade e pode ser utilizado em futuras análises que requerem a estrutura 3D do receptor. No entanto, como ainda existem alguns resíduos de aminoácidos em regiões não permitidas, o modelo foi refinado usando o servidor *GalaxyRefine*<sup>124</sup> para garantir que a estrutura tridimensional do receptor 5-HT<sub>6</sub> seja confiável. O *GalaxyRefine*<sup>124</sup> primeiro reconstrói todas as conformações das cadeias laterais, as re-empacota e posteriormente relaxa

a estrutura várias vezes por meio de simulações de *Dinâmica* molecular em curtos intervalos de tempo.<sup>124</sup>

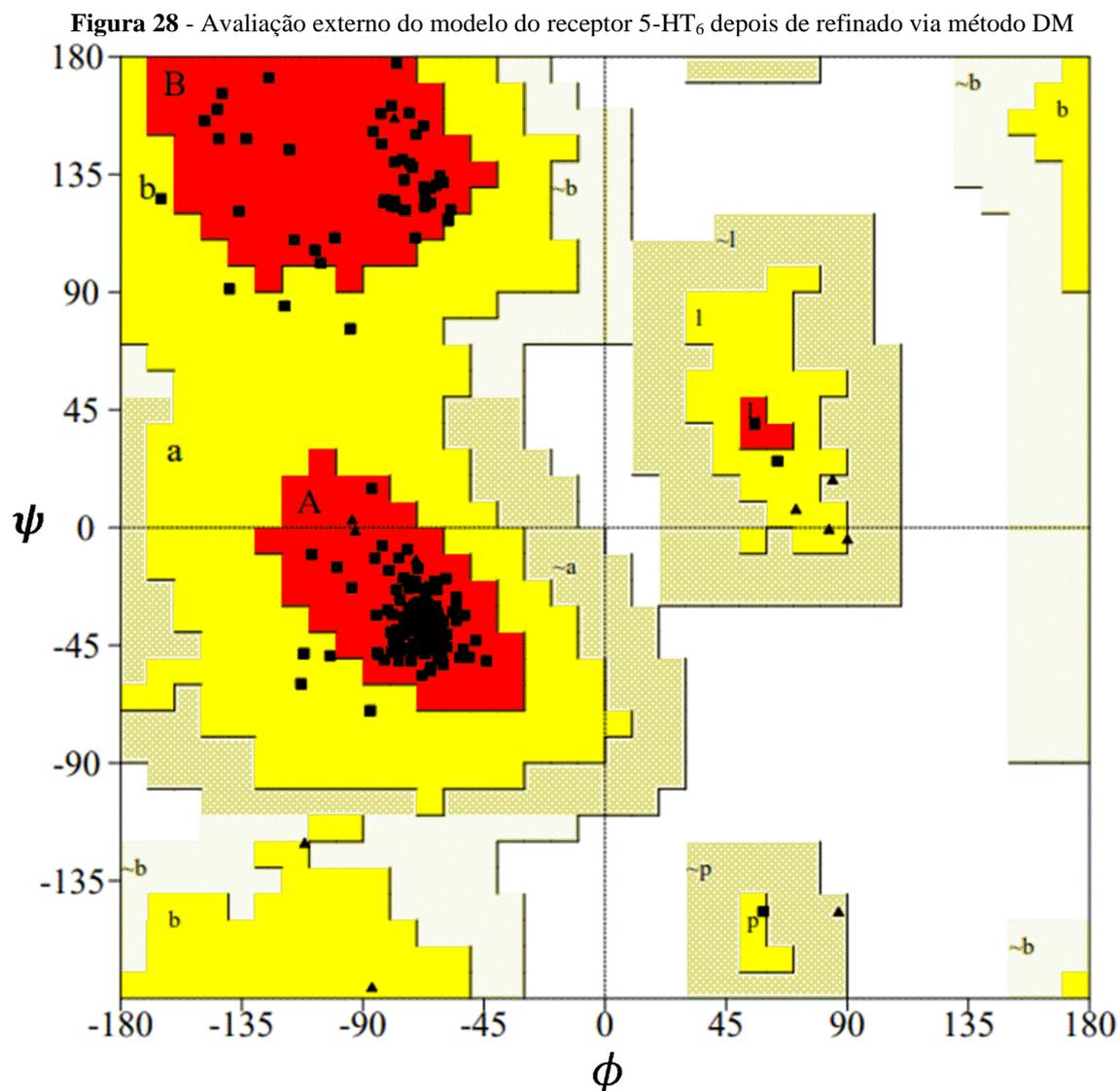
A Figura 27 mostra a diferença entre o primeiro modelo construído e o refinado, indicando que o modelo final realmente apresentou melhoras significativas. Essa comparação se deu por meio do cálculo do potencial estatístico denominado DOPE (do inglês, *Discrete Optimized Protein Energy*) que avalia a qualidade de um modelo 3D construído por método de predição. O potencial DOPE foi originado a partir de análises de várias proteínas com estruturas tridimensionais conhecidas, tendo, portanto, sua eficácia comprovada quando comparado a vários outros potenciais.<sup>154</sup>

Figura 27 – Comparação do modelo refinado por meio do cálculo da energia DOPE



Fonte: autoria própria

Para comprovar a performance do modelo final, outra avaliação externa foi realizada no servidor SAVES. Conforme visto na Figura 28, o resultado final mostra que o modelo refinado realmente é superior ao inicial.



**Fonte:** autoria própria

Os resultados da segunda análise da qualidade estereoquímica do modelo 5-HT<sub>6</sub> mostraram que 100% dos resíduos de aminoácidos se encontram em regiões favorecidas, dos quais 96,6% deles estão em regiões muito favorecidas e 3,4% em regiões permitidas. Portanto, como o modelo do receptor 5-HT<sub>6</sub> gerado passou por ambas as verificações interna e externa, o mesmo pode ser utilizado em futuros procedimentos que requerem sua estrutura tridimensional.

## 5.2 Geração dos modelos QSAR

Os modelos QSARs foram construídos com dois conjuntos de compostos (Tabelas 2 e 3), conforme resultados apresentados a seguir. Todos os compostos foram otimizados para que os descritores eletrônicos como, momento de dipolo, energia do ponto zero, polarizabilidade eletrônica, energia dos orbitais de fronteiras (HOMO e LUMO) dentre outros fossem obtidos. Além desses descritores, outros foram calculados, como os descritores topológicos, geométricos, constitucionais, índices de conectividade, etc. Ao total, obteve-se 1250 descritores para cada conjunto de dados. Em razão dessa imensa quantidade de descritores, uma pré-seleção foi realizada por meio do cálculo do coeficiente de correlação de Pearson, a fim de excluir aqueles menos correlacionados à atividade biológica.

### 5.2.1 Primeiro conjunto de compostos

Os descritores menos correlacionados ao  $pK_i$  foram excluídos, permanecendo 181. Estes foram submetidos à técnica de algoritmo genético para que aqueles mais relevantes fossem selecionados para a construção dos modelos QSAR-2D. A partir da técnica AG, foram gerados 100 modelos, variando-se o número de geração e o critério de corte (relacionado à correlação entre os descritores), dos quais, escolheu-se aquele que apresentou os melhores parâmetros estatísticos, conforme pontuado na Tabela 5.

**Tabela 5** - Parâmetros estatísticos analisados para escolher os descritores que participarão do modelo QSAR-2D

Descritores selecionados	R <sup>a</sup>	S <sup>b</sup>	F <sup>c</sup>	q <sup>2d</sup>
X5sol, BEHm3, MATS6p, MATS1v, RDF140m, GATS8e, E <sub>LUMO</sub>	0,90	0,3	34,47	0,75

<sup>a</sup>Coefficiente de correlação; <sup>b</sup>desvio padrão; <sup>c</sup>teste F; <sup>d</sup>coeficiente de correlação da validação cruzada.

Os valores dos descritores para cada composto do conjunto 1 (Tabela A1- Apêndice A) foram calculados na plataforma *E-dragon*<sup>155</sup> (exceção feita aos descritores eletrônicos). A Tabela 6 apresenta a classe e o tipo de descritor selecionado para a construção do modelo QSAR-2D.

**Tabela 6** - Definição dos descritores selecionados e classe para a qual eles pertencem

Descritor	Classe	Definição
X5sol	Topológico	Índice de Conectividade de Solvatação, chi-5.
BELm3	BCUT	Menor autovalor n.3 da Matriz de Burden/ponderada pela massa atômica.
MATS6p	2D autocorrelação	autocorrelação de Moran de lag 6 ponderada pela polarizabilidade.
MATS1v		autocorrelação de Moran de lag 1 ponderada pelo volume de van der Waals.
GATS8e		autocorrelação de Geary autocorrelação de lag 8 ponderado pela eletronegatividade de Sanderson.
RDF140m	RDF	Função de distribuição radial- 140/ ponderado pela massa.
ELUMO	Eletrônico	Energia do orbital LUMO

O modelo foi construído com 43 compostos no conjunto treino, validado internamente por meio dos procedimentos de validação cruzada e y-randomização e por último, procedimentos de verificação externa foram realizados para mensurar seu poder preditivo (para essa verificação, **16** compostos foram usados-Tabela 3). A Tabela 7 apresenta os parâmetros estatísticos obtidos a partir do método PLS, escolhendo-se o modelo que alcançou os melhores parâmetros (resultado em negrito -Tabela 7).

**Tabela 7** - Valores dos parâmetros estatísticos analisados para a seleção do melhor modelo

VL <sup>a</sup>	%Var <sup>b</sup>	SEV <sup>c</sup>	PRESS <sub>val</sub> <sup>d</sup>	SEC <sup>e</sup>	PRESS <sub>cal</sub> <sup>f</sup>	q <sup>2g</sup>	r <sup>2h</sup>
1	29,16	0,39	6,18	0,36	4,93	0,71	0,77
2	43,78	0,39	6,09	0,31	3,61	0,74	0,84
<b>3</b>	<b>49,60</b>	<b>0,39</b>	<b>5,97</b>	<b>0,31</b>	<b>3,51</b>	<b>0,74</b>	<b>0,85</b>
4	65,29	0,40	6,53	0,32	3,48	0,72	0,85
5	79,40	0,41	6,65	0,32	3,46	0,72	0,85
6	90,81	0,40	6,45	0,32	3,46	0,72	0,85
7	100	0,40	6,47	0,33	3,46	0,72	0,85

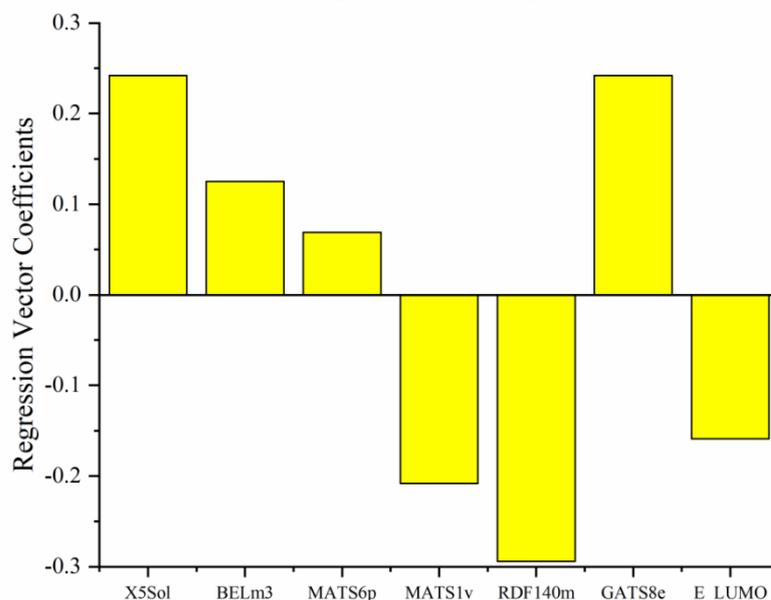
<sup>a</sup>Variável latente; <sup>b</sup>Variância percentual acumulada; <sup>c</sup>Erro padrão de validação; <sup>d</sup>Soma dos quadrados dos erros de previsão de validação; <sup>e</sup>Erro padrão de calibração; <sup>f</sup>Soma dos quadrados dos erros de previsão de calibração; <sup>g</sup>Coefficiente de validação cruzada; <sup>h</sup>Coefficiente de determinação múltipla.

De acordo com a literatura, modelos confiáveis devem apresentar  $q^2 > 0,5$  e  $r^2 > 0,6$ .<sup>156,157</sup> Com base nessa premissa, escolheu-se o modelo que atingiu os maiores valores de  $q^2$  e  $r^2$  e os menores valores de SEV, Press<sub>val</sub>, SEC e Press<sub>Cal</sub>.

O peso de cada descritor está representado na Figura 29 e observa-se que os descritores que apresentam valores positivos contribuem diretamente para a atividade biológica, isto é, em geral quanto maior o valor do descritor, maior a afinidade biológica da molécula. Mas vale salientar que, às vezes, os compostos mais ativos não se relacionam com os descritores que os descrevem de uma forma tão linear quanto se espera, uma vez que o peso dos demais descritores também influencia na resposta biológica. Dentre os descritores com coeficiente positivo, temos o X5Sol,

BELm3, MATS6p e GATS8e, já os descritores que apresentaram coeficiente negativo são MATS1v, RDF140 e ELUMO.

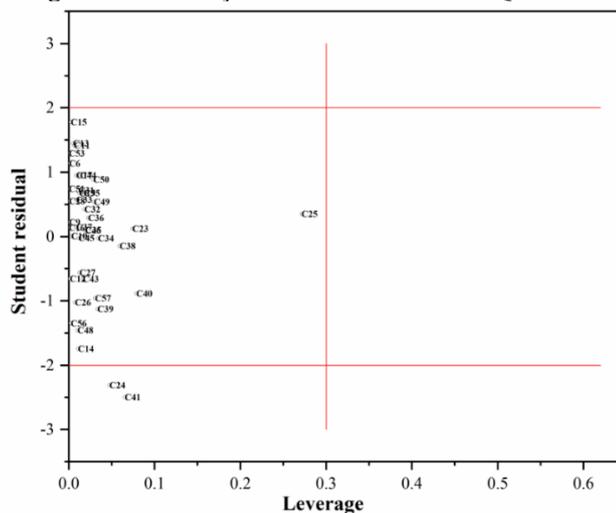
**Figura 29** - Contribuição de cada descritor para o vetor de regressão relacionado ao modelo PLS



Fonte: autoria própria<sup>136</sup>

Análises de detecção de *outlier* foram realizadas para garantir que o modelo não apresente amostras com comportamento discrepantes que possam influenciar em sua robustez. Essa análise é realizada por meio do gráfico de resíduos de *Student* (resultado da razão entre o resíduo de uma amostra e o seu desvio padrão) *versus leverage* (peso da influência de uma determinada amostra no modelo), como é mostrado na Figura 30. A linha vertical (vermelho) se refere aos valores de *leverage* e as linhas horizontais representam o intervalo em que os valores de resíduos de *Student* devem permanecer para serem considerados aceitáveis. Como mostra a Figura 30, apenas os compostos C24 e C41 apresentaram valores de resíduo de *Student* fora do limite requerido, no entanto esses compostos não foram considerados como amostras atípicas, pois os desvios observados são ínfimos e além do mais os valores de *leverage* estão dentro do limite considerado aceitáveis.<sup>158</sup>

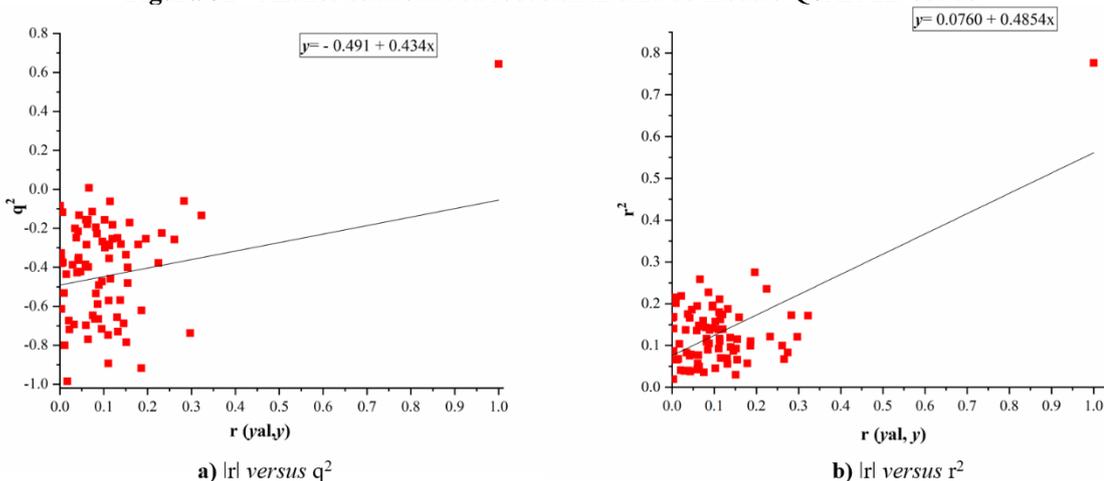
Figura 30 - Detecção de outlier no modelo QSAR-2D



Fonte: autoria própria<sup>136</sup>

Para certificar que o modelo não apresenta correlação ao acaso, dois gráficos foram construídos ( $r$  versus  $q^2$  e  $r$  versus  $r^2$ ), em que  $r$  é o coeficiente de correlação absoluto entre os valores de  $y$  original e  $y$  randomizado. Espera-se a partir desse procedimento que o valor do intercepto para ambos os gráficos seja menor que 0,05 e 0,3, respectivamente. Conforme mostra a Figura 31, esta condição foi atingida ( $r$  versus  $q^2 = -0,491$  e  $r$  versus  $r^2 = 0,0760$ ), atestando que o modelo possui boa consistência interna e portanto, é estatisticamente confiável.

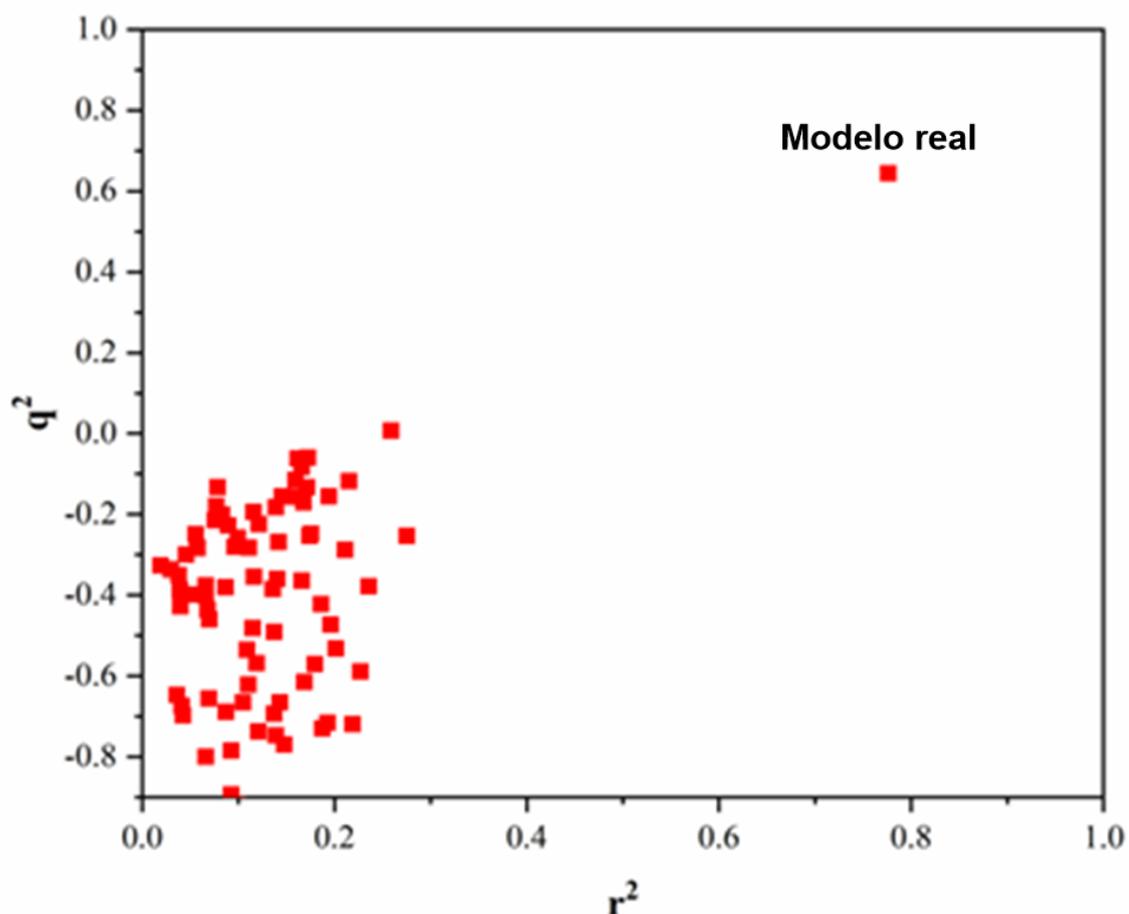
Figura 31 - Análise estatística da robustez interna do modelo QSAR-2D obtido



Fonte: autoria própria<sup>136</sup>

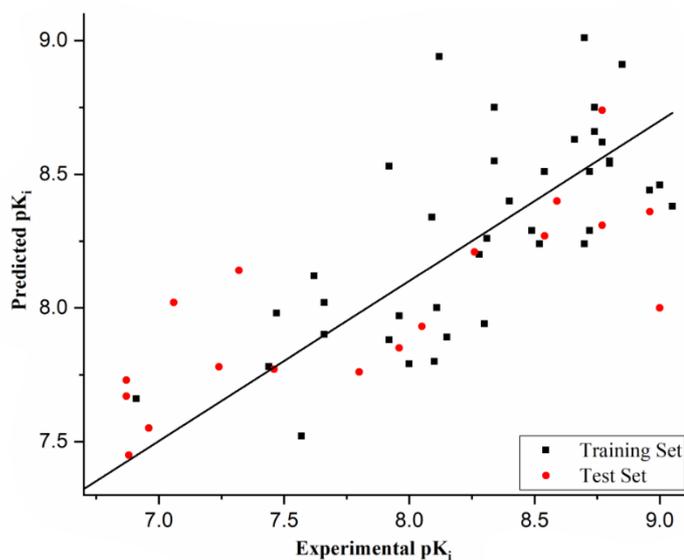
O teste de  $y$ -randomização foi realizado 80 vezes e ao final, espera-se que o modelo randomizado tenha baixa qualidade estatística, mostrando que o modelo real não foi obtido ao acaso. De fato, o modelo randomizado apresentou valores de  $q^2$  e  $r^2$  insignificantes (Figura 32).

Figura 32 - teste de y-randomização aplicado ao modelo



Fonte: autoria própria<sup>136</sup>

Além da validação interna, o modelo QSAR-2D foi avaliado externamente, em que foi obtido  $r^2_{\text{teste}} = 0,79$  (Figura 33) e portanto, a predição do efeito antagonista dos compostos do conjunto externo é mais uma prova do poder preditivo do presente modelo, mostrando que ele pode ser usado para prever valores de atividade biológica de potenciais antagonistas do receptor 5-HT<sub>6</sub>.

**Figura 33** - Predição dos valores de  $pK_i$  para os conjuntos treinamento e teste

Fonte: autoria própria<sup>136</sup>

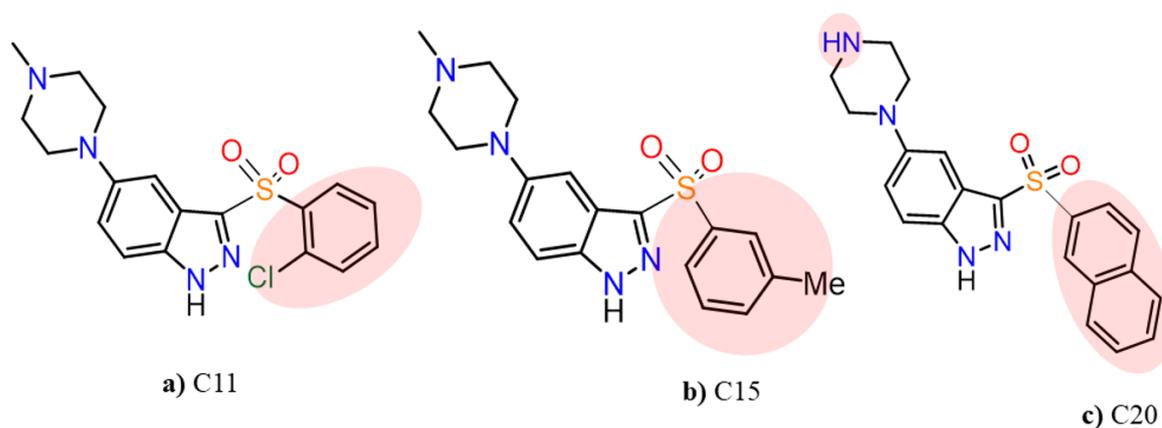
#### 5.2.1.1 Análise dos descritores selecionados para a construção do modelo QSAR-2D

A relação entre a estrutura molecular de um composto e seu valor de atividade biológica contribui para o entendimento das possíveis interações desses compostos com um dado alvo biológico, além de indicar quais grupos são mais favoráveis para aumentar determinado efeito biológico. Portanto, uma análise extremamente importante quando se emprega métodos QSARs é a compreensão do papel de cada descritor e como as informações extraídas deles explicam o efeito biológico dos compostos. As análises a seguir mostram como cada um dos descritores selecionados para a construção do modelo QSAR-2D se relaciona com o efeito antagonista dos compostos sob investigação.

**X5sol:** Índices de conectividade de solvatação estão relacionados à entropia de solvatação e mostram a dispersão no ambiente bioquímico.<sup>159</sup> Para o modelo, conforme já mostrado (Figura 29), o descritor X5sol tem contribuição positiva, isto é, aumentando o valor desse descritor, o valor de  $pK_i$  também aumenta. No entanto, a relação observada não foi tão linear quanto a esperada, pois conforme analisado, entre os três compostos mais ativos (C15, C11 e C20,  $pK_i$  igual a 9,05; 9,00 e 8,05, respectivamente - Figura 34), aquele com o menor valor de  $pK_i$  (C20) apresentou o maior valor para esse descritor (X5sol = 9,45). A diferença mais acentuada entre esses compostos está na região destacada em vermelho (Figura 34), para a qual percebeu-se que quanto maior a hidrofobicidade nessa região (maior a entropia no ambiente bioquímico), maior

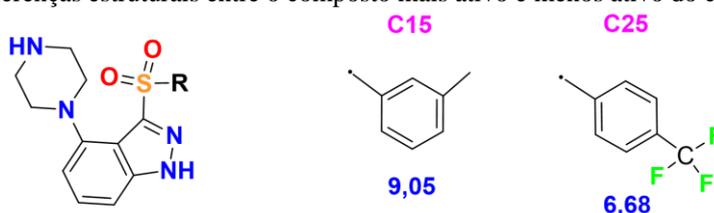
o valor desse descritor. No entanto, parece que há um certo limite, pois há grupos muito volumosos nessa região com características hidrofóbicas que não contribuem para o efeito biológico esperado. O composto **C25** apresenta grupo volumoso demais na referida região (Figura 35), apesar de manter as características hidrofóbicas necessárias, mas apresenta o menor valor de atividade biológica ( $pK_i = 6,68$ ), mostrando que é preciso ter ponderação ao se adicionar grupos hidrofóbicos, pois ao mesmo tempo que esses grupos contribuem para aumentar a resposta biológicas, se eles forem demasiadamente volumosos, o efeito biológico diminui.

**Figura 34** - Estrutura molecular dos três compostos mais ativos do conjunto de dados



Fonte: autoria própria<sup>136</sup>

**BELm3**: este tipo de descritor pertence à classe de descritores BCTU, desenvolvidos por Burden<sup>160</sup>, em que os parâmetros de Burden são organizados em uma matriz que inclui informações a respeito do número atômico de cada átomo e o tipo de ligação para os átomos ligados e não ligados. Além disso, descritores dessa classe descrevem a geometria dos compostos e codificam várias propriedades em um único valor. O descritor BELm3 é ponderado pela massa atômica, tendo contribuição positiva para a atividade biológica e conforme observou-se, os compostos com maiores valores de massa molecular detinham os maiores valores de atividade biológica. Contudo, grupos muito volumosos na posição **R** com maiores graus de liberdade não são tolerados nesta posição, por exemplo, os compostos **C15** e **C25** apresentam o menor e maior grupo nesta posição, e o maior e menor valor de  $pK_i$ , respectivamente (Figura 35).

**Figura 35** - Diferenças estruturais entre o composto mais ativo e menos ativo do conjunto de dados

Fonte: autoria própria

**2D Autocorrelação (MATS6p, MATS1v e GATS8e):** os descritores MATS6p, MATS1v e GATS8e são ponderados pela polarizabilidade, volume de van der Waals e eletronegatividade, respectivamente. O primeiro e o último descritor contribuem positivamente para o efeito biológico, ao passo que o segundo possui uma relação inversa. Tais descritores podem ser obtidos a partir da aplicação da seguinte equação (Eq.15)<sup>161</sup>:

$$AC = \int_a^b f(x) \cdot f(x+l) dx \quad 15$$

Em que  $f(x)$  é relacionado a qualquer propriedade físico-química para cada átomo da molécula e o conjunto de átomos são representados como pontos discretos no espaço e a propriedade atômica (por exemplo, polarizabilidade, volume de van der Waals e eletronegatividade) é uma função de avaliação desses pontos.

Em geral, os maiores valores do descritor MATS6p foram atribuídos aos compostos mais ativos, o mesmo foi observado para o descritor GATS8e que está relacionado a eletronegatividade, e de fato, essa propriedade contribui muito para o entendimento da interação entre alguns ligantes e o receptor 5-HT<sub>6</sub>, conforme será mostrado na Seção 5.4. Em relação ao descritor MATS1V, os menores valores desse descritor estão associados aos compostos mais ativos, e como ele está relacionado ao volume de van der Waals, observou-se mais uma vez que os compostos com substituintes menores na posição R (Figuras 34 e 35) tem ótimos valores de pK<sub>i</sub> e por isso podem interagir mais eficazmente no sítio ativo do receptor 5-HT<sub>6</sub>.

**RDF140m:** a função de distribuição radial de um composto pode ser entendida como uma distribuição de probabilidade de encontrar um átomo em um volume esférico de raio  $r$ . A equação 16 mostra como o descritor RDF é obtido<sup>161</sup>:

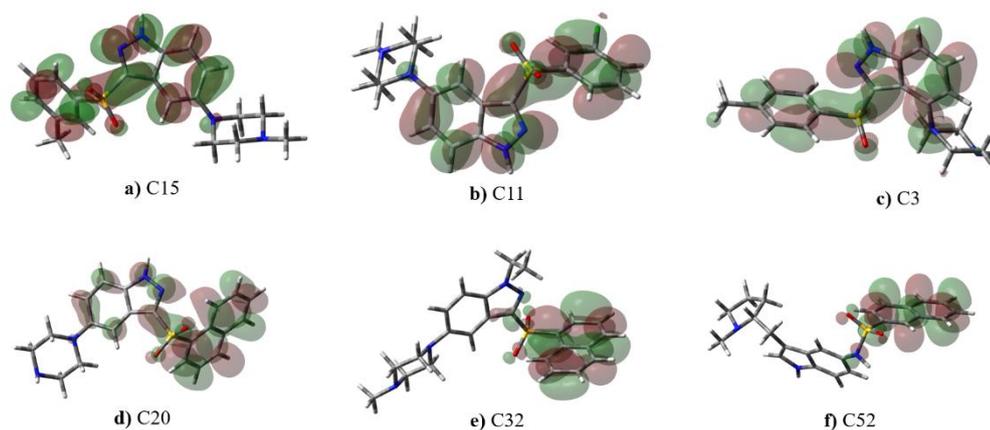
$$g(r) = f \sum_i^{N-1} \sum_{j>1}^N A_i A_j e^{-B(r-r_{ij})^2}$$

16

Onde  $N$  representa o número de átomos;  $f$  é um fator de escala;  $A$  é a propriedade atômica do átomo  $i$  e  $j$ ;  $r_{ij}$  é a distância entre os átomos  $i$  e  $j$ ; e  $B$  é um parâmetro de suavização (matematicamente, esse parâmetro assegura que a função terá um bom comportamento). A propriedade atômica de interesse para a presente análise se trata da massa atômica dos átomos  $i$  e  $j$  e os maiores valores desse descritor foram assinalados aos compostos menos ativos. A explicação para isso, novamente, está relacionada ao tamanho do grupo R ligado ao substituinte sulfonil (Tabela 2).

**ELUMO:** Sabe-se que os menores valores da energia LUMO são atribuídos a compostos com maior capacidade de aceitar elétrons<sup>162-165</sup>, sendo assim, analisou-se os orbitais LUMOs dos compostos mais ativos para averiguar as regiões com maior densidade eletrônica. As análises indicam que as regiões mostradas na Figura 36 podem ter papel destacado ao interagir com o receptor 5-HT<sub>6</sub> e importantes informações podem ser extraídas, o que de fato ocorreu, conforme discutido na Seção 5.4.

**Figura 36** - Orbitais LUMOs dos compostos mais ativos



**Fonte:** autoria própria<sup>136</sup>

## 5.2.1.2 Análises QSAR-3D

A fim de entender as características estéricas e eletrostáticas dos compostos que influenciam à propriedade biológica, um modelo QSAR-3D foi gerado usando o método CoMFA. Inicialmente, esse modelo foi construído usando os parâmetros padrões do *software* Sybyl 8.2 (Tripos, Inc)<sup>149</sup>, em que foi selecionada a opção de focagem de região de maneira a refinar e melhorar estatisticamente o modelo. De fato, ocorreu uma melhora significativa no resultado final do modelo, como pode ser comprovado pelos parâmetros estatísticos mostrados na Tabela 8.

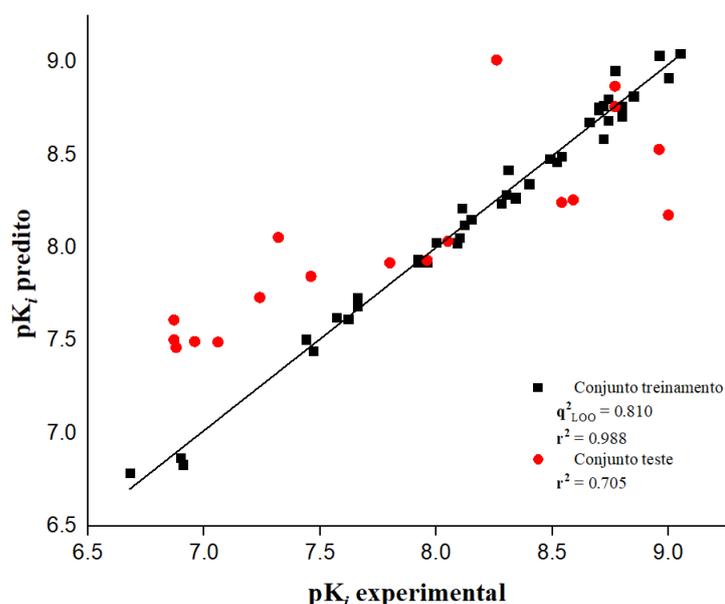
**Tabela 8** - Resultados estatísticos para os modelos “sem focagem” e melhor focagem encontrada

Parâmetro estatístico	Sem focagem	$w = 0.5 / d = 1.0 \text{ \AA}$
$q^2_{\text{LOO}}$	0,449	0,810
$r^2$	0,976	0,988
SEE	0,099	0,072
SEP	0,451	0,280
E	0,478	0,564
S	0,522	0,436
N	2,000	6,000

$q^2_{\text{LOO}}$  = Coeficiente de validação usando o método leave-one-out; **SEP** = Erro padrão de predição; **N** = Número de componentes principais geradas a partir do método PLS;  $r^2$  = Coeficiente de predição; **SEE** = Erro padrão de estimação; **S** = Contribuição dos campos estéricos; **E** = contribuição dos campos eletrostáticos; **w** = peso de desvio padrão; **d** = Fator grid de distância.

Os resultados apresentados na Tabela 8 mostram que o melhor modelo obtido, utilizando as técnicas de focagem de região, foi o que possui um espaçamento de 1,0 Å e o peso de 0,5 entre os pontos do *grid*, além dos seguintes parâmetros estatísticos:  $q^2_{\text{LOO}} = \mathbf{0,810}$ ; **SEP = 0,280**; **PCs = 6**;  $r^2 = \mathbf{0,988}$ ; **SEE = 0,280**; **S = 0,430**; **E = 0,564**. Essa análise demonstra a importância de utilizar as metodologias de focagem, já que houve melhora significativa na qualidade estatística do modelo, e os resultados indicam que o modelo obtido apresenta consistência interna aceitável e que não há diferença apreciável entre as contribuições dos campos estereoquímicos e eletrostáticos. A Figura 37 apresenta o gráfico dos valores de  $\text{pK}_i$  experimentais *versus* preditos a partir da metodologia CoMFA para os compostos dos conjuntos de treinamento e teste.

**Figura 37-** Valores de  $pK_i$  experimentais *versus* preditos obtidos a partir do modelo CoMFA



Fonte: autoria própria<sup>136</sup>

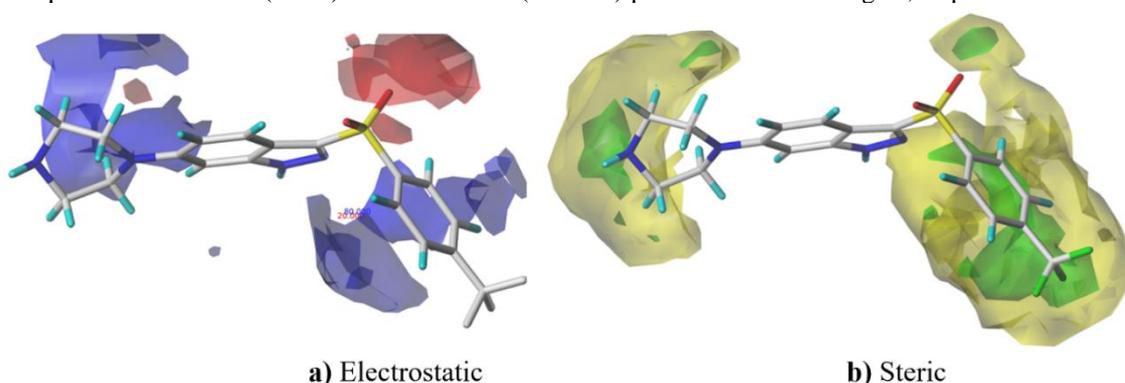
Embora não se tenha como comparar os resultados, pois metodologias diferentes foram utilizadas para a construção dos dois modelos (QSARs-2D e 3D), os compostos selecionados para o conjunto teste são os mesmos para ambos os modelos. Conforme apresentado na Tabela 9, os valores da validação externa ( $r^2_{pred} = 0,705$ ) mostram ótima concordância entre os valores de  $pK_i$  experimentais e preditos, e, por isso os valores de resíduos também são pequenos, atestando que o modelo QSAR-3D gerado é robusto e tem boa capacidade preditiva.

**Tabela 9** - Resíduos calculados para os compostos do conjunto teste

Composto	$pK_i$ experimental	$pK_i$ predito	Resíduo
C1	6,87	7,504	0,6342
C3	6,87	7,611	0,7405
C4	7,06	7,492	0,4324
C5	7,32	8,056	0,7355
C7	8,26	9,013	0,7526
C8	8,77	8,761	-0,0094
C18	8,59	8,258	-0,332
C20	8,96	8,531	-0,4291
C21	7,96	7,933	-0,0271
C22	6,96	7,495	0,5352
C29	8,54	8,245	-0,2952
C30	8,77	8,872	0,1018
C42	7,80	7,919	0,1187
C47	8,05	8,035	-0,0148
C52	9,00	8,177	-0,8232
C54	7,46	7,846	0,3858
C58	7,24	7,732	0,4922
C59	6,88	7,462	0,5818

Após as análises estatísticas das validações externas, confirmando a boa capacidade preditiva do melhor modelo CoMFA obtido, foram visualizados os mapas de contorno 3D. Esses mapas permitem a visualização das regiões com maiores contribuições estereoquímicas e eletrostáticas, as quais podem lançar luz para o entendimento das diferenças entre os valores de afinidade biológica das estruturas moleculares do conjunto de dados (Tabela 2). Os mapas foram construídos para três compostos, escolhidos com base em seus valores de atividade biológica, a saber: menos ativo (**C25**), efeito antagonista moderado (**C47**) e o mais ativo (**C15**), conforme visto nas Figuras 38 a 40.

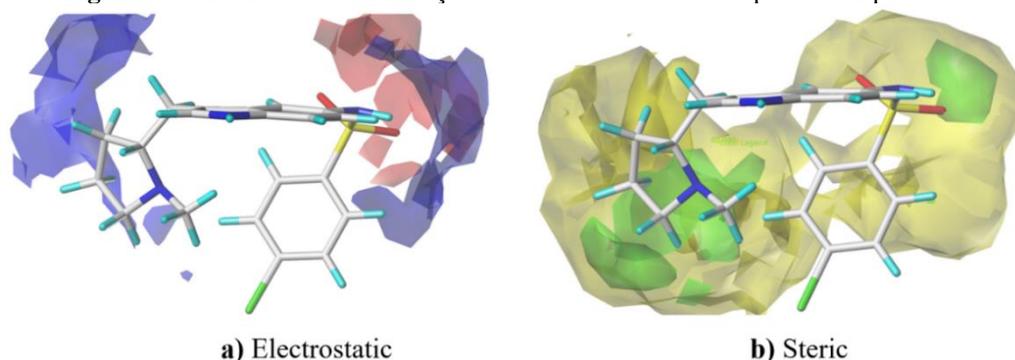
**Figura 38** - Análises das contribuições eletrostáticas e estéricas para o composto 25. **a)** contribuições eletrostáticas positivas (azul) e negativas (vermelho) para a atividade biológica, respectivamente; **b)** contribuições estereoquímicas favoráveis (verde) e desfavoráveis (amarelo) para a atividade biológica, respectivamente



Fonte: autoria própria<sup>136</sup>

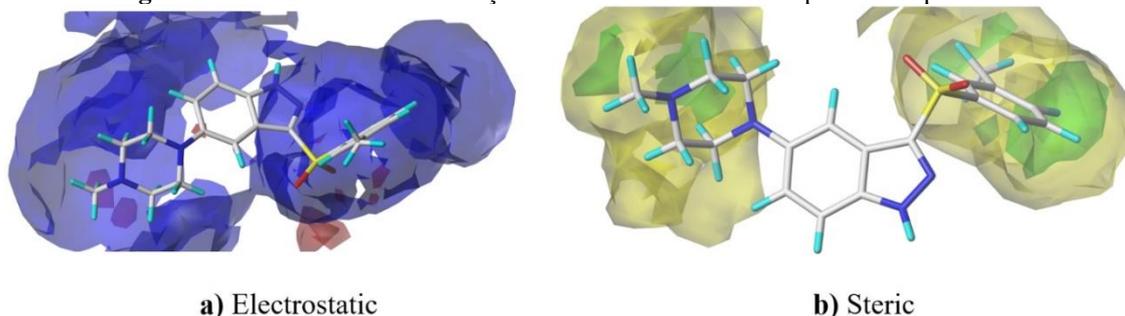
Percebe-se que há uma competição entre as contribuições favoráveis e desfavoráveis (contribuições estereoquímicas – Figura 38b) na região HYD (substituinte trifluormetil-benzeno) do composto **C25**. Análises anteriores (Seção 5.2.1.1) apontaram que substituintes com características hidrofóbicos são bem tolerados nessa região, no entanto o fator volume influencia negativamente na resposta biológica. Na região PI percebe-se a importância de grupos com átomos ionizáveis, como por exemplo o nitrogênio da piperazina que apresentou contribuição favorável para a atividade biológica.

Já para contribuições eletrostáticas positivas, os dois anéis das extremidades apresentaram-se favoráveis para a atividade biológica, e os poliedros próximos ao grupo sulfonil apresentam contribuições eletrostáticas negativas, sugerindo substituições por um grupo carregado eletrostaticamente negativo.

**Figura 39** - Análises das contribuições eletrostáticas e estéricas para o composto 47

Fonte: autoria própria<sup>136</sup>

A região vermelha próxima ao átomo de nitrogênio e ao enxofre sugere que substituições por grupos eletronegativos podem contribuir favoravelmente para a atividade biológica. A região próxima ao nitrogênio 2 da piperazina sugere uma significativa contribuição estérica, contudo, a conformação do composto mostra um grande impedimento estérico, uma vez que os dois anéis das extremidades, piperazina e benzeno, estão bem próximos, aumentando o volume da região.

**Figura 40** - Análises das contribuições eletrostáticas e estéricas para o composto 15

Fonte: autoria própria<sup>136</sup>

A grande área em azul indica que substituições por grupos eletropositivos podem favorecer a atividade biológica. Contudo, nas extremidades existem 2 anéis aromáticos que também apresentam impedimento estérico. Devido à ressonância dos anéis aromáticos, há essa grande contribuição favorável eletrostaticamente (por grupos eletropositivos). A região com o sulfonil apresentou contribuições eletrostáticas negativas para a molécula.

## 5.2.2 Segundo conjunto de compostos

A primeira etapa para a construção do modelo ANN- QSAR seguiu o mesmo protocolo do implementado para a geração do modelo QSAR-2D (Seção 5.2.1). 181 descritores permaneceram após a seleção dos descritores via coeficiente de correlação de Pearson, número ainda considerado grande e por isso, a técnica de algoritmo genético foi empregada para selecionar os descritores mais relevantes para a construção do modelo. Para essa tarefa, foram gerados 200 modelos variando alguns parâmetros estatísticos, tais como: número de geração ( $G_i$  – 600 a 1000), número de descritores (5 a 9) e critério de correlação entre os descritores. A partir daí, escolheu-se os seguintes descritores (Tabela 10):

**Tabela 10** - Parâmetros estatísticos analisados para a seleção dos descritores

$G_i$	Critério	Seleção dos descritores	$R^a$	$S^b$	$F^c$	$q^{2d}$
600	0,9	JhetZ, BEHm2, GATS3p, GATS1p, RDF025u, Au, H3v, $E_{HOMO(-2)}$	0,77	0,45	13,13	0,52

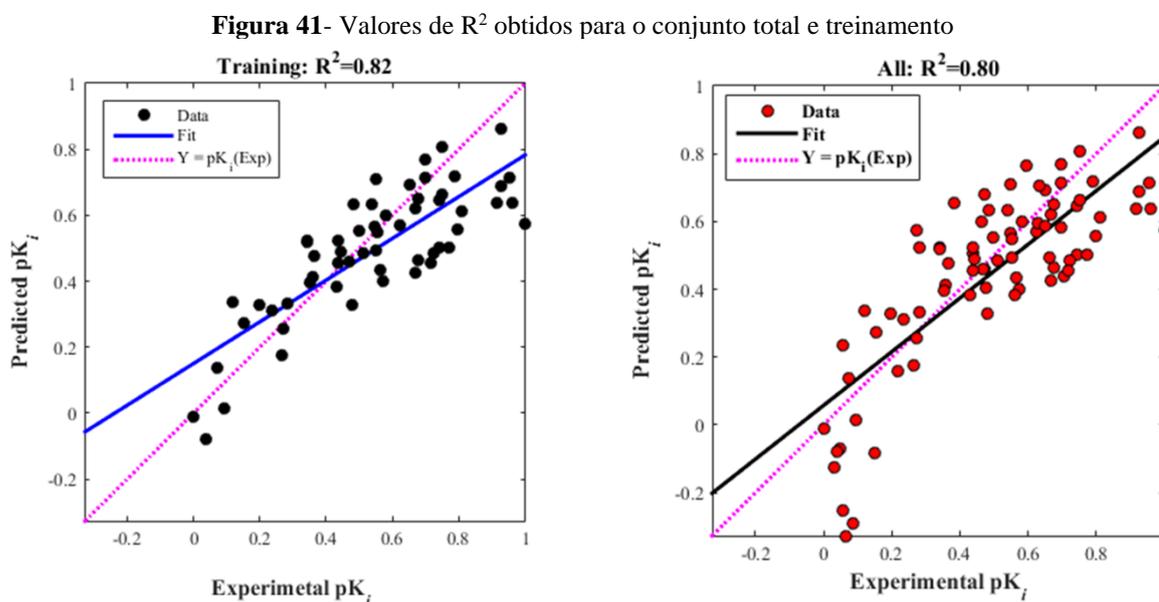
<sup>a</sup>Coefficiente de determinação múltipla; <sup>b</sup>desvio padrão; <sup>c</sup>teste F; <sup>d</sup>coeficiente de validação cruzada.

A geração igual a **600** indica o número de vezes que o AG buscou pelos melhores descritores, o critério **0,9** mostra que os descritores mais correlacionados entre si foram excluídos, permanecendo apenas um deles para a geração do modelo. A Tabela 11 apresenta os descritores selecionados, a classe e a definição de cada um (os valores dos descritores para cada composto do conjunto 2 estão no Apêndice B - Tabela B1).

**Tabela 11** - Definição dos descritores selecionados para a construção do modelo ANN -QSAR

Descritor	Classe	Definição
JhetZ	Topológico	Índice do tipo Balaban a partir da Matrix Z/ponderado pela distância de matrix
BEHm2	BCUT (Burden- CAS University of Texas Eigen values)	Maior autovalor da matrix de Burden de n-2 / ponderado pela massa atômica.
GATS3p	2D Autocorrelation	Autocorrelação de Geary - lag 3/ponderado pela polarizabilidade atômica.
GATS1p		Autocorrelação de Geary -lag 1/ponderado pela polarizabilidade atômica.
RDF025u	RDF (Função de distribuição radial)	Função de distribuição radial - 25/unweighted.
Au	WHIN (Weighted Holistic Invariant Molecular)	A total size index/unweighted
H3v	GETAWAY (Geometric Topology and Atom Weights Assembly)	Autocorrelação de H de lag3/ponderado pelo volume atômico van der Waals.
$E_{HOMO(-2)}$	Eletrônico	Energia do orbital $HOMO_{(-2)}$

Para a construção do modelo ANN-QSAR, foi usado a rede do tipo *Perceptron* multicamada (do inglês, *Multilayer Perceptron* -MLP) com a arquitetura 8-20-1 (8 neurônios na camada de entrada, 20 na camada intermediária, 1 na camada de saída). As funções de transferências que apresentaram os melhores resultados foram *logsig-tansig* e o algoritmo de treinamento selecionado foi o Levenberg-Marquardt (LM). O próprio algoritmo divide a rede em três conjuntos: treinamento (70%), teste (20%) e validação (10%), além disso, um modelo com todos os compostos também foi gerado para que seus valores de  $R^2$  sejam comparados aos  $R^2$  dos demais conjuntos. Conforme mostra a Figura 41, ambos os valores de  $R^2$  estão próximos e, portanto, os requisitos de boas práticas em QSAR foram atendidos.<sup>158</sup>



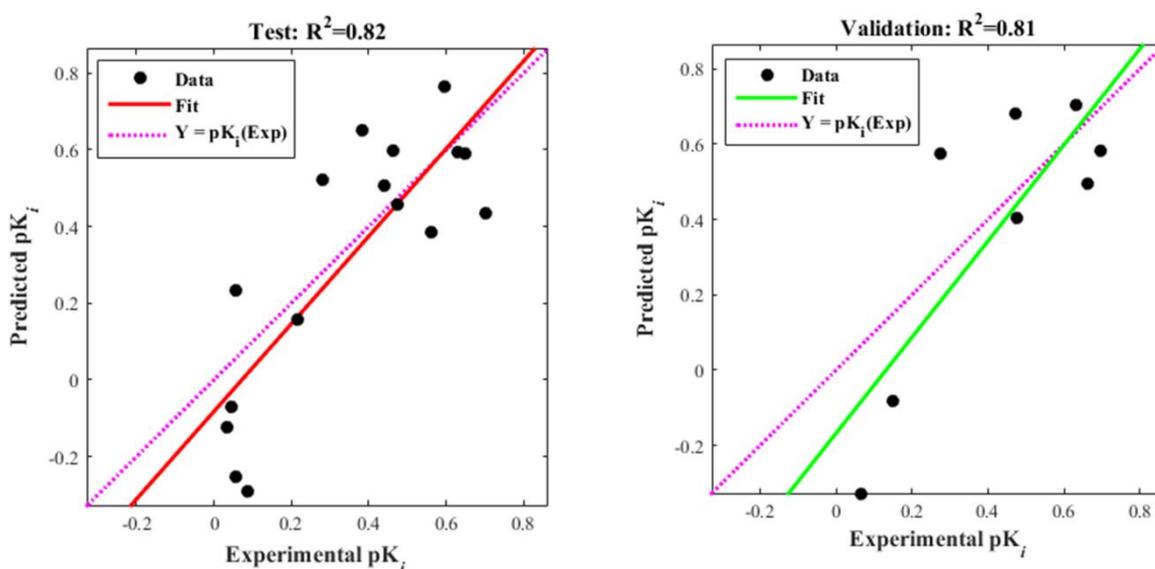
Fonte: autoria própria<sup>137</sup>

A Figura 41 mostra que os modelos gerados a partir dos conjuntos treinamento e total explicam 82% e 80%, respectivamente, das informações originais de ambos os modelos. A linha pontilhada (magenta) representa o resultado ideal, em que os valores de  $pK_i$  preditos são iguais aos valores experimentais, já a linha azul (conjunto treinamento) e preta (conjunto total) representa o melhor ajuste encontrado pelo modelo ANN-QSAR.

Para garantir a robustez do modelo, o poder preditivo e a qualidade de ajuste do modelo, dois outros conjuntos foram gerados (Figura 42), os quais apresentaram valores de  $R^2_{\text{teste}} = 0,82$  e  $R^2_{\text{validação}} = 0,81$ . É importante observar que ambos os valores de  $R^2$  (conjunto total, treino, teste e validação- Figuras 41 e 42) permaneceram próximos um do outro, provando que o modelo é consistente e, portanto, pode ser utilizado para prever valores de  $pK_i$  de novas amostras. O conjunto teste é usado como uma prova robusta que certifica a capacidade preditiva

do modelo ANN-QSAR e o conjunto validação é necessário para garantir que não haja superajuste na rede neural. Entende-se ainda que este conjunto pode agir como um procedimento de validação cruzada, uma vez que os erros obtidos são usados para estimar a capacidade preditiva do modelo, bem como sua consistência interna. Além do mais, uma vez que se obtenha erros exacerbados, a rede pode ser retreinada de modo a se obter melhores resultados.

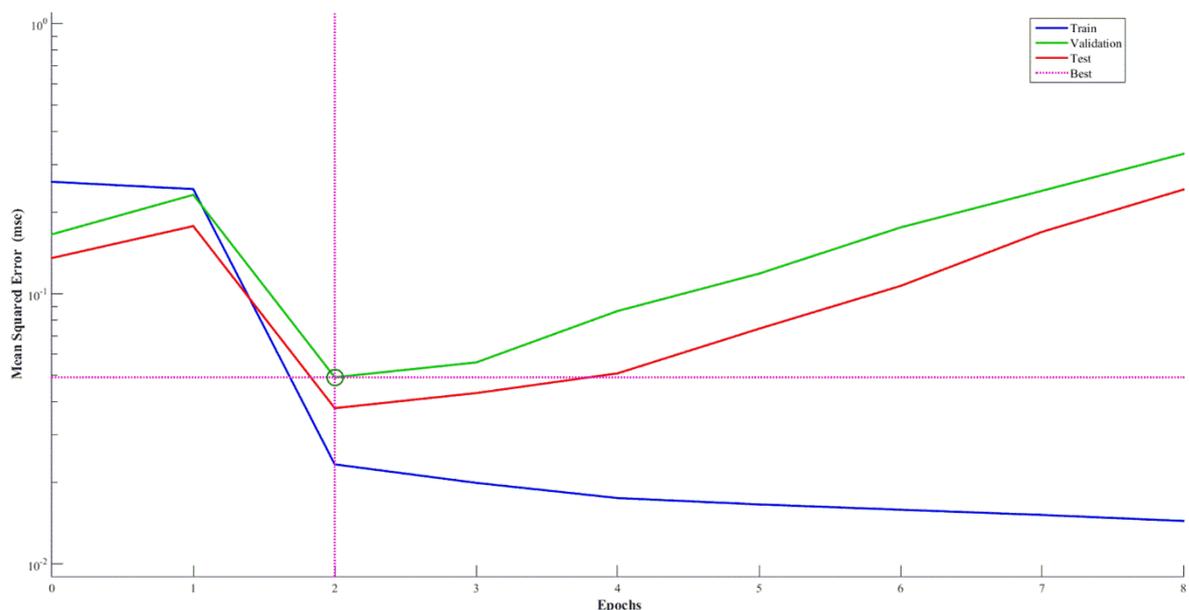
**Figura 42** - Valores de  $R^2$  obtidos para o conjunto teste e validação



Fonte: autoria própria<sup>137</sup>

A performance neural foi avaliada por meio da análise do MSE (Erro Quadrado Médio, do inglês, *Mean-Squared Error*) para o conjunto treinamento, teste e validação (Figura 43). observa-se ainda que havendo um aumento significativo no MSE do conjunto validação, o algoritmo LM trava automaticamente o treinamento da rede, indicando que a rede deve ser retreinada para se obter resultados confiáveis.

**Figura 43** - Erro quadrado médio obtido para os conjuntos treinamento, teste e validação em função do número de iterações



Fonte: autoria própria<sup>137</sup>

O ponto verde mostra a melhor performance da rede neural, em que se obteve o menor MSE. Como esperado, para que um modelo ANN seja considerado satisfatório, o valor de MSE para o conjunto treinamento deve estar abaixo dos valores de erros dos demais conjuntos, além do mais, se a curva de erro do conjunto teste for consideravelmente maior do que a do conjunto validação, então é possível que haja superajuste no modelo. Provando a confiabilidade do modelo ANN-QSAR, esses achados não foram atingidos, conforme observa-se na Figura 43.

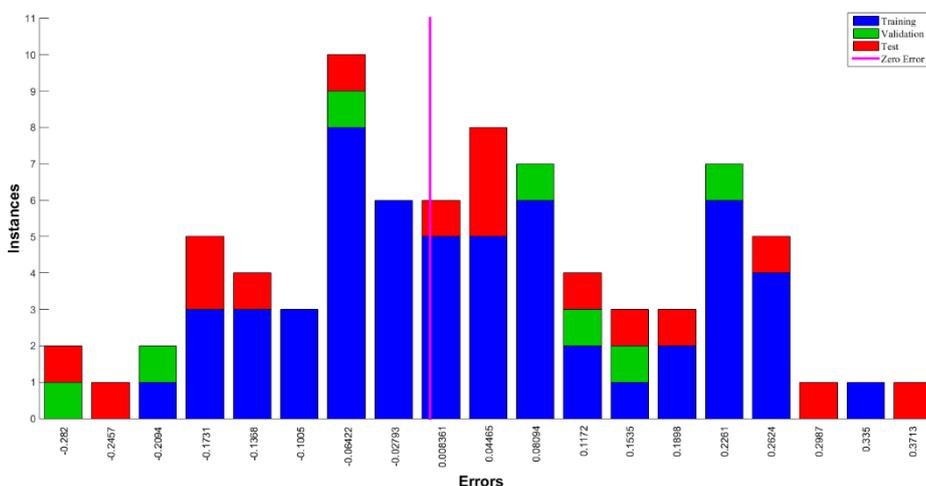
A Tabela 12 contém os valores de  $pK_i$  preditos para todos os compostos do conjunto teste e os erros associados, em que a grande maioria dos compostos exibiram erros menores que 5%, mostrando mais uma vez que o modelo gerado possui boa capacidade preditiva.

Tabela 12 - Valores de  $pK_i$  preditos para todos os compostos do conjunto teste

Composto	$pK_i$ experimental	$pK_i$ predito	Resíduo absoluto	Resíduo relativo%
11	7,53	7,52	0,01	0,13
21	7,04	7,28	-0,24	3,41
25	7,51	7,64	-0,13	1,73
27	7,76	7,58	0,18	2,32
28	7,45	7,51	-0,06	0,81
35	7,94	7,90	0,04	0,50
36	7,84	8,01	-0,17	2,17
42	7,99	7,93	0,06	0,75
52	7,30	7,57	-0,27	3,70
68	8,12	7,86	0,26	3,20
73	6,44	6,32	0,12	1,86
75	6,88	6,82	0,06	0,87
76	6,46	6,64	-0,18	2,79
77	6,40	6,24	0,16	2,50
78	6,54	6,16	0,38	5,81
81	6,46	6,15	0,31	4,80

Uma outra análise considerada essencial, é a detecção de amostras com comportamentos anômalos. Como resultado, observa-se a partir da Figura 44 que nenhuma amostra apresentou esse comportamento, sendo esse um teste adicional para certificar que o modelo apresenta consistência interna, e, portanto, confiabilidade.

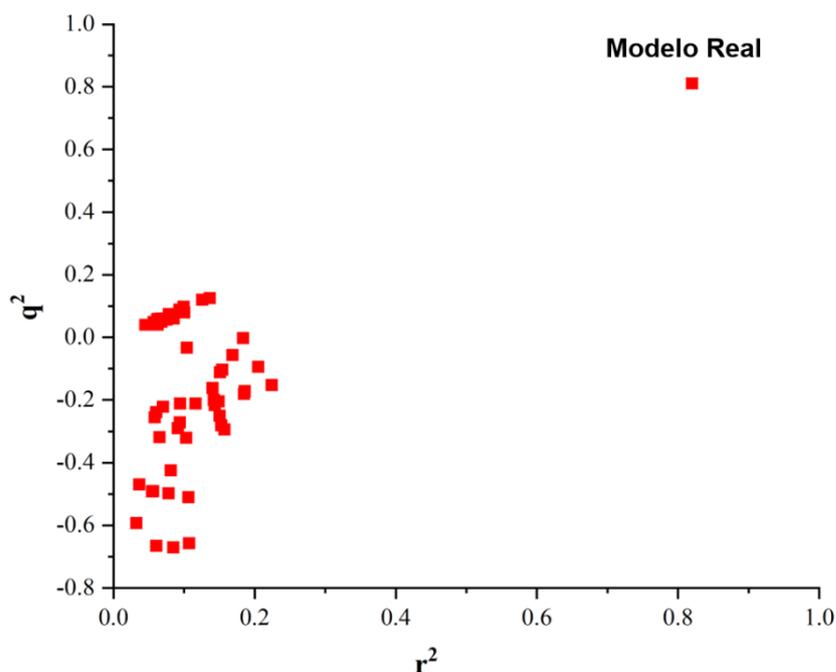
Figura 44 - Detecção de outlier para todos os compostos do conjunto 2



Fonte: autoria própria<sup>137</sup>

O procedimento de y-randomização também foi realizado, no qual as amostras foram randomizadas 80 vezes, além de teste de *leave-one-out* para cada repetição. Os resultados para esse teste mostram que os valores de  $q^2$  e  $r^2$  são insignificantes (Figura 45) diante do modelo real, provando que o modelo ANN- QSAR não foi obtido ao acaso.

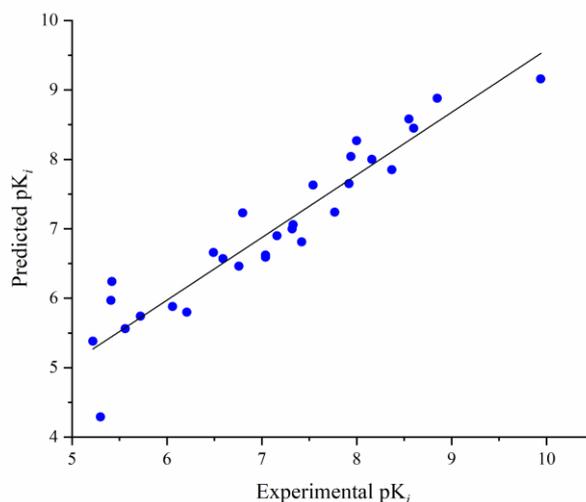
**Figura 45** - Verificação da consistência interna do modelo ANN-QSAR



Fonte: autoria própria<sup>137</sup>

Após ser submetido a vários procedimentos estatísticos, O modelo ANN-QSAR foi usado para prever a atividade biológica de vários antagonistas do receptor 5-HT<sub>6</sub>, selecionados a partir do banco de dados ChEMBL, certificando o poder preditivo desse modelo. Os resultados obtidos são bem próximos aos valores de  $pK_i$  experimental, conforme mostrado na Figura 46 e Tabela 13.

**Figura 46** - Valores de  $pK_i$  experimental *versus* preditos pelo modelo ANN-QSAR para os compostos selecionados do ChEMBL



Fonte: autoria própria<sup>137</sup>

Salienta-se que o modelo ANN-QSAR construído obteve valor de  $pK_i$  para o composto **215945** (código ChEMBL – Tabela 13) igual ao experimental, sendo esse teste uma prova indubitável da capacidade preditiva do modelo diante de moléculas com atividade antagonista perante ao receptor 5-HT<sub>6</sub>.

Ressalta-se ainda que os compostos (**411**, **808**, **402143**, **2159463**, **2159456** e **3260806** – Tabela 13), cujas características estruturais diferem das propostas por López-Rodríguez e colaboradores<sup>82</sup> (Figura 13) possuem baixos valores de  $pK_i$  e o modelo ANN-QSAR foi capaz de prever os seus valores de  $pK_i$  dentro de uma pequena margem de erro, conforme mostra a Tabela 13.

**Tabela 13-** Valores de  $pK_i$  preditos pelo modelo ANN-QSAR

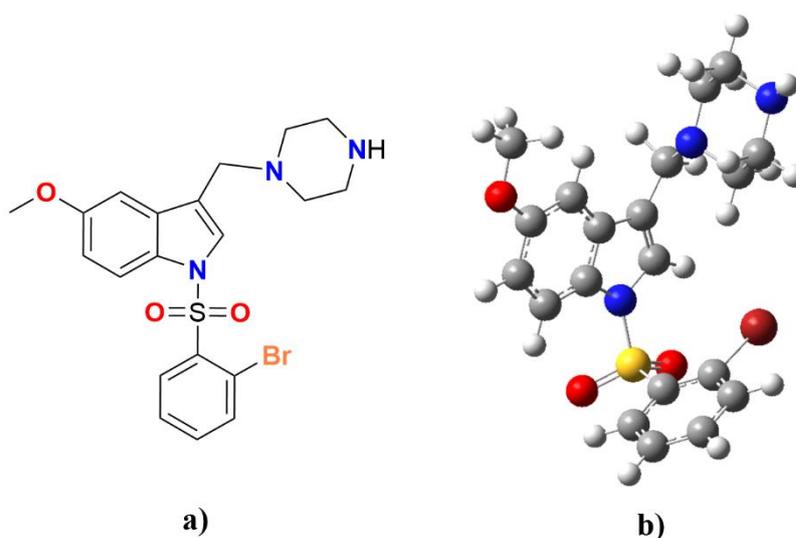
<b>Código ChEMBL</b>	<b><math>pK_i</math> experimental</b>	<b><math>pK_i</math> predito</b>	<b>Resíduo absoluto</b>
411	5,22	5,38	-0,16
808	5,30	4,29	1,01
62193	6,21	5,80	0,41
24474	9,94	9,16	0,78
402143	5,41	5,97	-0,56
484164	8,55	8,58	-0,03
482990	8,37	7,85	0,52
506736	8,16	8,00	0,16
483559	7,94	8,04	-0,10
482748	7,54	7,63	-0,09
347335	6,80	7,23	-0,43
2159470	6,59	6,57	0,02
2159463	5,42	6,24	-0,81
2159464	6,06	5,88	0,19
2159481	6,49	6,66	-0,17
<b>2159456</b>	<b>5,56</b>	<b>5,56</b>	<b>0,00</b>
154641	8,60	8,45	0,15
46071	7,92	7,65	0,27
3260792	7,32	7,00	0,32
3260806	5,72	5,74	-0,02
1834222	7,04	6,62	0,42
1834351	7,04	6,59	0,45
3260790	7,16	6,90	0,26
60669	8,00	8,27	-0,27
95136	8,85	8,88	-0,03
434365	7,42	6,81	0,61
378656	7,77	7,24	0,53
413049	6,76	6,46	0,30
433461	7,33	7,06	0,27

### 5.2.3 Análise dos descritores selecionados

Os descritores selecionados são baseados na dimensionalidade das estruturas, por exemplo 2D e 3D (Figura 47). Os descritores 2D (JhetZ, BEHm2, GATS3p e GATS1p) são

representados por um único valor, obtido a partir das representações 2D dos compostos, conforme visto na Figura 47. A partir desses descritores, é possível extrair importantes informações por exemplo, tamanho, grau de ramificação, flexibilidade, vizinhos de átomos como nitrogênio, oxigênio e bromo (e a partir daí, é possível intuir sobre os efeitos estéricos e eletrostáticos, por exemplo). Por outro lado, os descritores 3D (Au, RDF025u, H3v e HOMO<sub>(-2)</sub>) podem fornecer informações sobre o reconhecimento molecular entre os compostos e o alvo molecular.

**Figura 47** - Dimensionalidade das representações estruturais - a) descritores 2D e b) descritores 3D



Fonte: autoria própria<sup>137</sup>

**Descritores topológicos (JhetZ):** o JhetZ é um descritor do tipo Balaban codificado como uma matriz de distância, cujas distâncias são computadas entre os pares<sup>165</sup>, conforme exemplo abaixo (Figura 48).

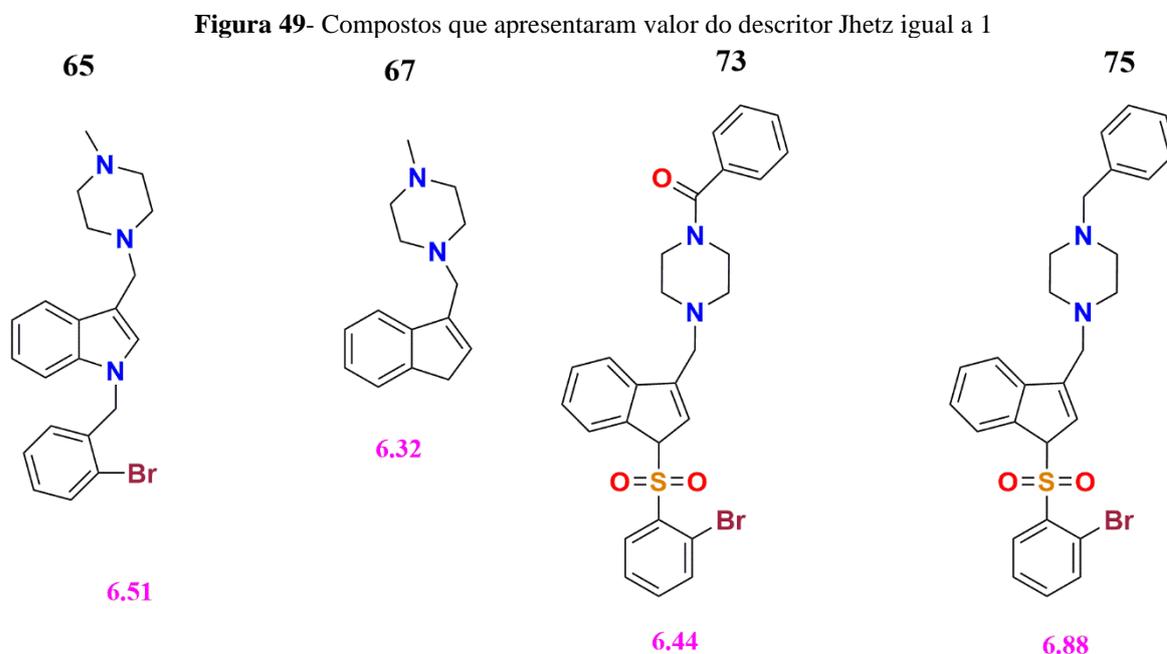
**Figura 48**- exemplo de uma matriz de distância para a molécula isopentano

Átomo	1	2	3	4	5
1	0	1	2	3	2
2	1	0	1	2	1
3	2	1	0	1	2
4	3	2	1	0	3
5	2	1	2	3	0

Fonte: autoria própria

Esse descritor transforma as estruturas moleculares em um grafo molecular (Figura 48) e é considerado altamente discriminatório, uma vez que ele aponta a existência de heteroátomos e

ligações múltiplas entre os átomos de um dado composto.<sup>165</sup> Para a série de compostos sob estudo (conjunto 2 – Tabela 3), a maioria das moléculas apresentaram valor do descritor JhetZ igual a 2, exceto os compostos **65**, **67**, **73** e **75** (Figura 49), os quais exibiram valor igual a 1.



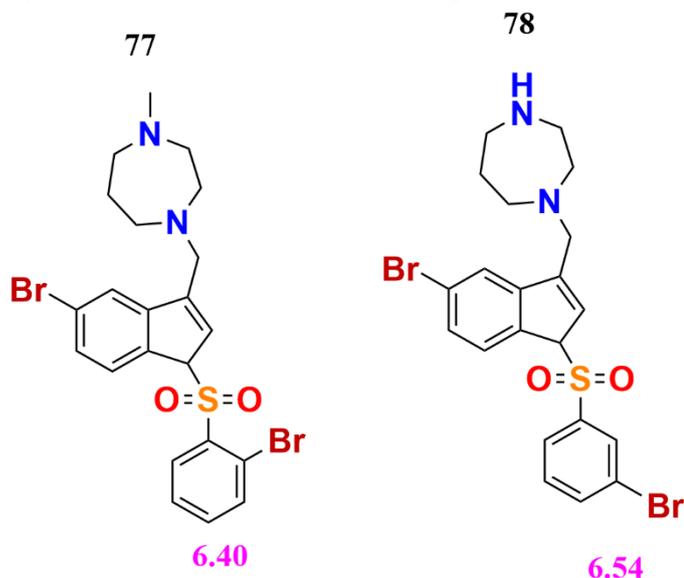
Fonte: autoria própria<sup>137</sup>

A partir da análise desse descritor, algumas características vieram à luz:

- Todos os compostos que apresentaram em suas estruturas moleculares mais que três anéis aromáticos exibiram índice de Balaban igual a 1;
- Os compostos que não possuem em suas estruturas grupos sulfonil ou grupos com características similares também exibiram índice de Balaban igual a 1 (Compostos **65** e **67**), indicando que esses tipos de grupos podem ter relevância considerável no processo de reconhecimento molecular;
- Moléculas, cujas estruturas apresentam substituintes acetofenona e tolueno (compostos **73** e **75**, respectivamente) ligados ao grupo PI (ver Figura 13), têm índices de Balaban também igual a 1 e valores de atividade biológica abaixo de 7, indicando que, talvez, esses tipos de anéis aromáticos não contribuem significativamente para aumentar o efeito antagonista dos compostos.

**Descritor BCTU (BEHm2):** Em geral, os maiores valores desse descritor (Tabela **B1** – Apêndice **B**) foram atribuídos aos compostos com maiores valores de atividade biológica, com exceção dos compostos **77** e **78** (Figura **50**), cujo comportamento foi contrário ao observado.

**Figura 50-** Compostos com valor do descritor BEHm2 igual a 6, mas com baixo valor de  $pK_i$



Fonte: autoria própria<sup>137</sup>

Algumas características da relação desse descritor com os compostos incluem: **a)** grupos volumosos na posição PI (ver Figura **13**) não contribuem para aumentar os valores de  $pK_i$  das moléculas; **b)** halogênios ligados ao grupo AR (ver Figura **13**) não são tolerados (compostos **77** e **78**- Figura **50**) e, portanto, não impulsionam o aumento do efeito biológico esperado; **c)** em geral, *o*-bromobenzeno ou *m*-bromobenzeno não parecem contribuir para o efeito antagonista dos compostos quando há átomos de halogênios na Região AR.

**Autocorrelação 2D (GATS3p e GATS1p):** Uma discussão sobre as características dessa classe de descritores já foi realizada na Seção **5.2.1.1**, por isso a abordagem aqui será focada apenas nos descritores do tipo GATS (GATS3p e GATS1p)<sup>165</sup>, os quais matematicamente podem ser definidos como (Eq.17):

$$c_k = \frac{\frac{1}{2\Delta_k} \sum_{i=1}^A \sum_{j=1}^A (w_i - w_j)^2 \cdot \delta(d_{ij}; k)}{\frac{1}{(A-1)} \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad 17$$

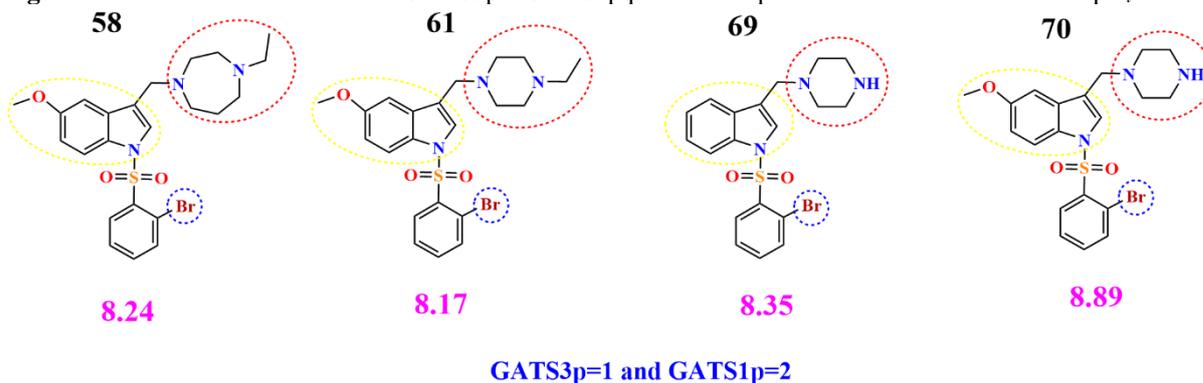
Em que  $w_i$  representa alguma propriedade eletrônica (polarizabilidade para esse estudo),  $\bar{w}$  se trata do valor médio dos valores de polarizabilidade,  $A$  refere-se ao número atômico,  $d_{ij}$  é a distância topológica entre os átomos  $i$  e  $j$ ,  $\delta(d_{ij}; k)$  representa o delta de Kronecker:

$$\left\{ \begin{array}{l} \delta_{i,j} = 1 \text{ if } d_{i,j} = k \\ \delta_{i,j} = 0 \text{ if } d_{i,j} \neq k \end{array} \right\}, \Delta_k \text{ simboliza o número de pares de átomos para distâncias igual a } k.$$

A faixa de valores entre 0 e 1 indica autocorrelações positivas, já valores maiores que 1 apontam uma autocorrelação inversa. Traduzindo para nosso estudo, os átomos das moléculas representam um conjunto de pontos discretos em um espaço 3D, no qual a função relacionada à propriedade atômica (polarizabilidade) é computada a partir desses pontos. Então, ambos os descritores GATS3p e GATS1p referem-se à autocorrelação de polarizabilidade em intervalos de distâncias igual a 3 e 1, respectivamente.

Todos os compostos exibiram valores igual ou maior que 1 para ambos os descritores (GATS3p e GATS1p – Tabela B1- Apêndice B). Para o descritor GATS3p não há autocorrelação nos intervalos de distância topológica igual a 3, enquanto para o descritor GATS1p a autocorrelação é inversa. De fato, os compostos com os maiores valores de  $pK_i$  possuem átomos de halogênios (Br – halogênio com maior poder polarizante entre os existentes nas estruturas do conjunto 2- Tabela 3) nos substituintes presentes na região HYD (ver Figura 13). Portanto, a polarizabilidade nessa região é igual para os compostos 58, 61, 69 e 70 (região destacada em azul na Figura 51) ao passo que nas regiões PI e AR desses compostos, há diferentes grupos polarizantes (regiões amarela e vermelha- Figura 51).

Figura 51 - Análises dos descritores GATS3p e GATS1p para os compostos com maiores valores de  $pK_i$

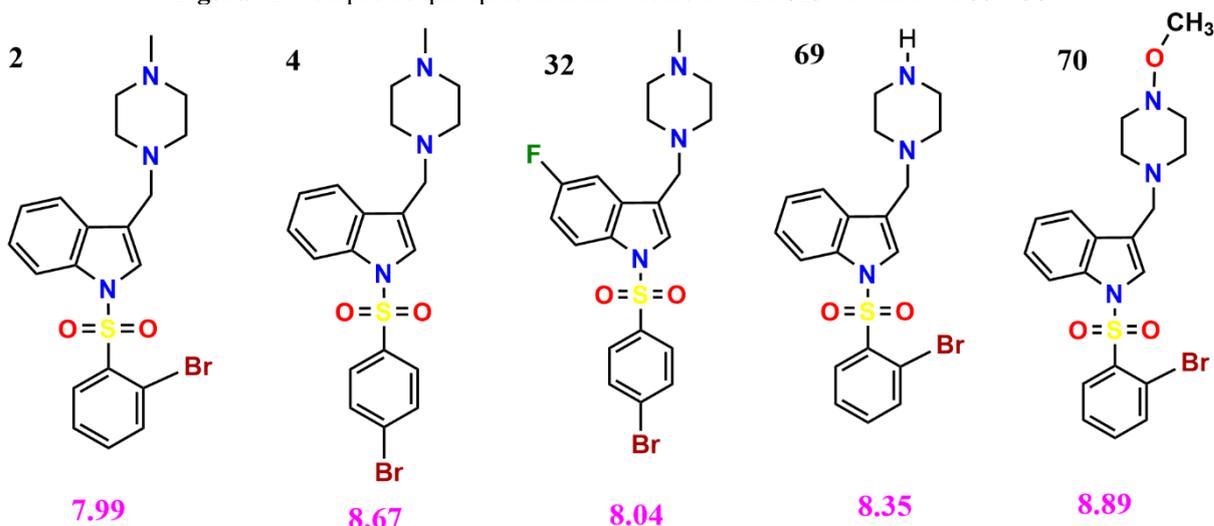


Fonte: autoria própria<sup>137</sup>

**RDF025u:** para os compostos estudados neste conjunto (Tabela 3), os valores do descritor RDF025u são 28, 35 a 50, 52 e 54, e a partir daí observou-se que o menor valor desse descritor (28) foi computado para o composto menos ativo (composto 67). Já os compostos com valores

de RDF025u na faixa de **35** a **38** possuem pelo menos um átomo de halogênio em suas estruturas químicas (exceção feita aos compostos **68** e **80**). Destaca-se os compostos **2**, **4**, **32**, **69** e **70** que exibiram valores de  $pK_i$  maior que **7,99** e átomos de Bromo ligados na posição 2' ou 4' dos substituintes na região HYD, indicando que talvez a presença desse átomo nessa região específica pode ser a razão do aumento do efeito antagonista dos compostos (Figura 52).

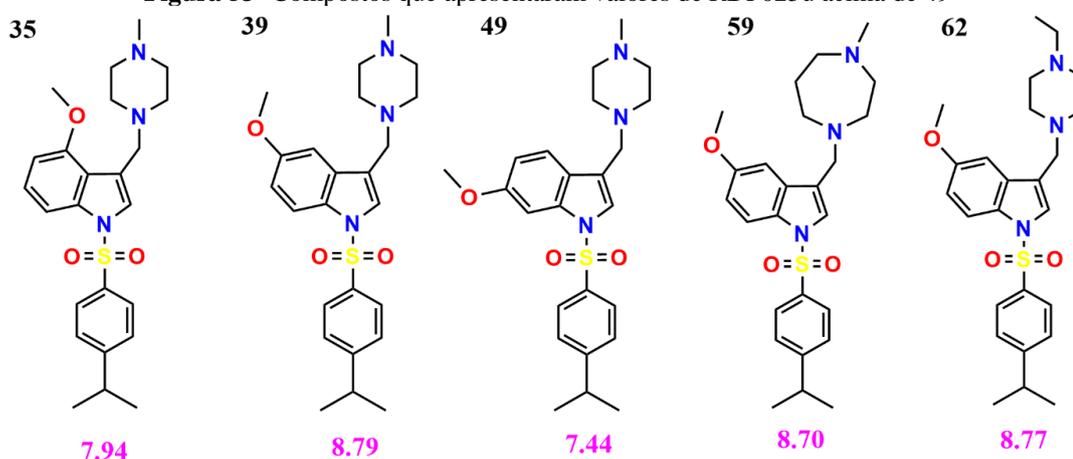
**Figura 52-** compostos que apresentaram valores de RDF025u na faixa de 35 a 38



Fonte: autoria própria<sup>137</sup>

No caso das moléculas com valores de RDF25u acima de **49** (compostos **35**, **39**, **49**, **59** e **62**) possuem substituinte metóxi ligados na posição 4', 5', ou 6' e grupos isopropil na posição 4' da região HYD. Esses compostos estão entre aqueles com maior valor de  $pK_i$ , então, intui-se que esses substituintes são bem tolerados nessa região e contribuem para aumentar o efeito inibitório dos compostos.

**Figura 53-** Compostos que apresentaram valores de RDF025u acima de 49



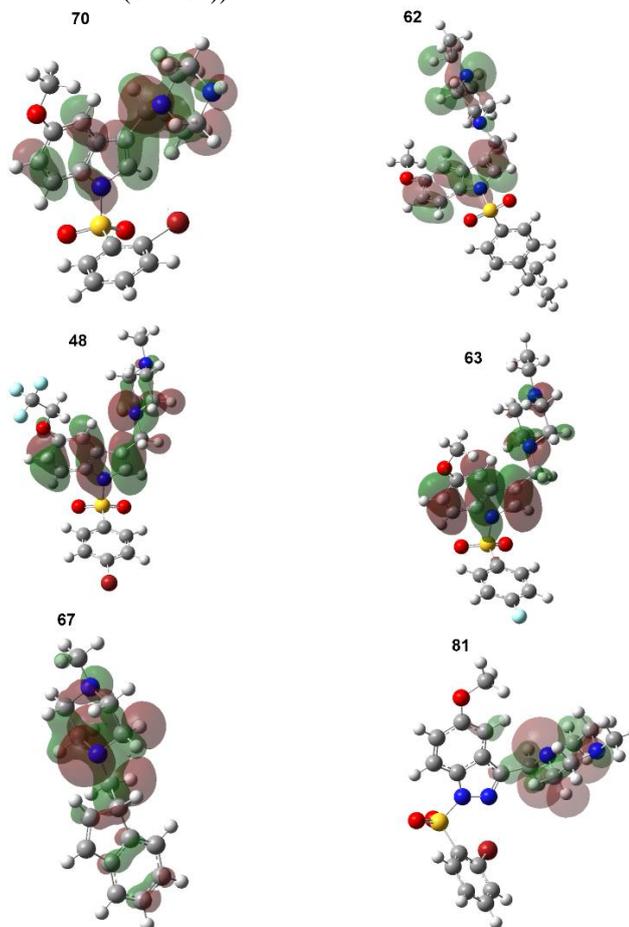
Fonte: autoria própria<sup>137</sup>

**AU**: esse tipo de descritor fornece uma dimensão global sobre as características de uma molécula e os autovalores desse descritor estão relacionados ao tamanho da molécula.<sup>164</sup> Observou-se a partir da análise deste descritor que o menor valor de AU (**33**) foi atribuído ao composto com menor potencial inibitório ( $pK_i = 6,32$ ) que também apresenta o menor volume em comparação aos demais compostos do conjunto (Tabela 3).

**H3v**: esse descritor é ponderado pelo volume de van der Waals e é computado a partir de uma matriz de *leverage*, cujos valores são obtidos a partir das coordenadas atômicas.<sup>167</sup> Para os presentes compostos, temos valores de H3v igual a 1 ou 0, sendo que a maioria deles possuem valor igual a 1, exceção feita para os compostos **67** e **80** que tiveram valor igual a 0. Esses compostos estão entre os menos ativos do grupo e possuem os menores volumes. Outro ponto a salientar, nenhum desses compostos possuem átomos de halogênios em suas estruturas, átomos esses que quando introduzidos em regiões específicas da molécula (região HYD) parecem contribuir positivamente para aumentar o efeito antagonista dos compostos.

**E<sub>HOMO(-2)</sub>**: esse descritor eletrônico representa a energia do orbital HOMO para uma molécula e indica sua capacidade de doar elétrons.<sup>164</sup> A presença desse descritor se justifica, uma vez que o processo de transferência eletrônica ocorre provavelmente na região PI. De fato, a presença de átomos de Nitrogênio nessa região comprova essa premissa conforme mostra a Figura 52, em que a representação dos orbitais HOMO é exibida para as moléculas mais ativas (compostos **62** e **70**), as que apresentam valores de  $pK_i$  moderados (compostos **63** e **48**) e as menos ativas (compostos **67** e **81**).

**Figura 54** - representação dos orbitais HOMO para os compostos mais ativos (**62** e **70**), os com efeito biológico moderado (**63** e **48**) e os menos ativos (**67** e **81**)



Fonte: autoria própria<sup>137</sup>

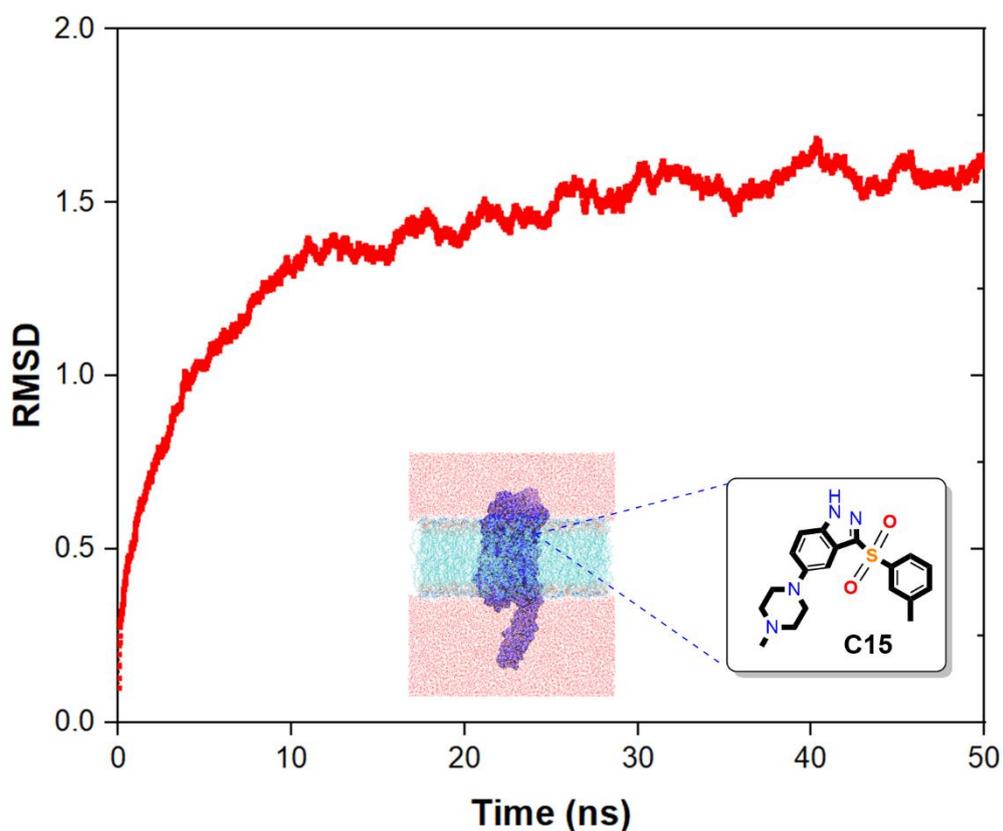
### 5.3 Análises da simulação de Dinâmica molecular

Realizou-se 50 ns de simulação de *Dinâmica* molecular, para a qual escolheu-se o composto mais ativo do conjunto 1 (**C15**) complexado ao receptor 5-HT<sub>6</sub>, o qual foi embebido em seu ambiente fisiológico (receptor, membrana, íons, solvente e ligante). Desse tempo, 20 ns foi usado para se atingir a estabilidade do sistema.

Um dos parâmetros analisados quando se deseja inferir algo acerca da estabilidade do sistema é o RMSD (do inglês, *root-mean-square deviation*, desvio médio quadrado). O sistema é considerado estável quando poucas flutuações são observadas ao longo de uma determinada faixa de tempo. A partir da Figura **55**, percebe-se que o sistema flutua até próximo 1,5 de RMSD, depois disso flutuações mínimas também são observadas, mas estas são desconsideradas pois não contribuem para alterar as propriedades termodinâmicas do sistema, indicando que o sistema apresenta baixas transições conformacionais. Uma vez que o receptor

5-HT<sub>6</sub> não possui estrutura cristalográfica conhecida, as simulações por DM foram essenciais para proceder com as análises receptor-ligante *via* Docking molecular com razoável grau de confiabilidade, conforme discutido na Seção 5.4.

**Figura 55** - Tempo de simulação *versus* RMSD para o receptor 5-HT<sub>6</sub> complexado ao ligante C15 embebida em seu ambiente fisiológico

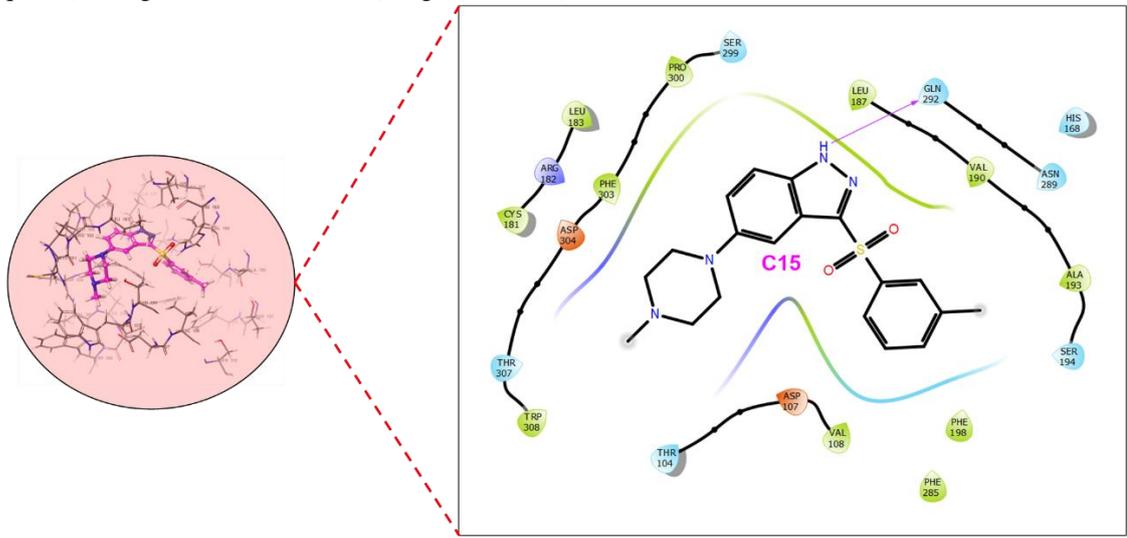


Fonte: autoria própria

#### 5.4 Análises das simulações de Docking molecular

Para realizar a simulação de *Docking* molecular, selecionou-se o composto mais ativo (**C15**) e o menos ativo (**C25**) do conjunto 1 (Tabela 2) para avaliar o modo de ligação e as diferenças entre esses compostos. A Figura 56 mostra os principais tipos de interações entre o **C15** e o sítio de ligação do receptor 5-HT<sub>6</sub>.

**Figura 56** - Principais tipos de interações entre o **C15** e o sítio ativo do receptor 5-HT<sub>6</sub>: representação 3D (imagem à esquerda) e diagrama bidimensional (imagem à direita).

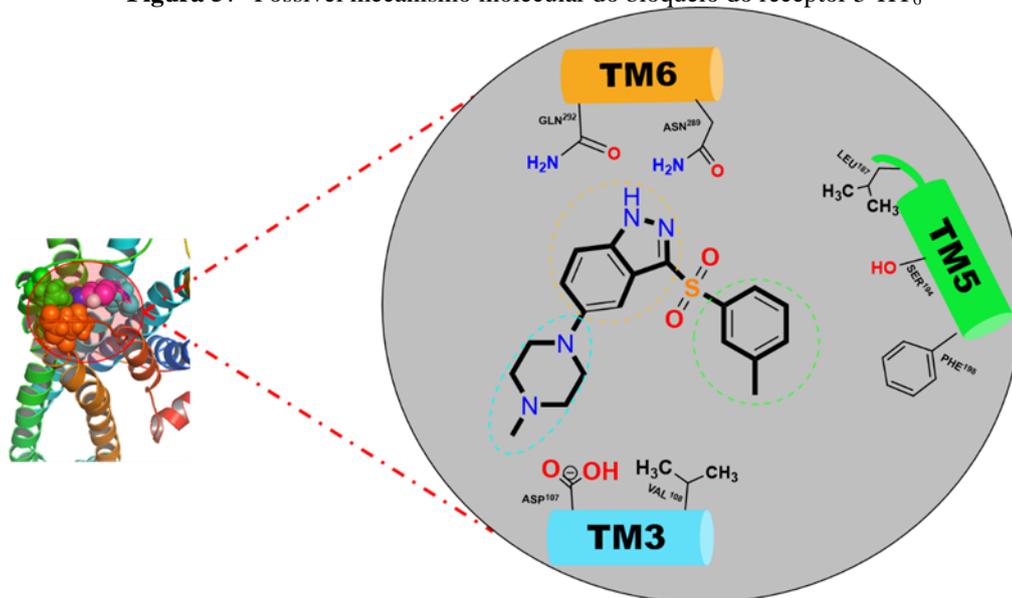


- Resíduos polares;
- Resíduos hidrofóbicos;
- Resíduos carregados negativamente;
- Resíduos carregados positivamente.

Fonte: autoria própria

A Figura 57 mostra as principais interações entre os resíduos de aminoácidos e o **C15**, bem como a localização desses resíduos no receptor.

**Figura 57**- Possível mecanismo molecular do bloqueio do receptor 5-HT<sub>6</sub>



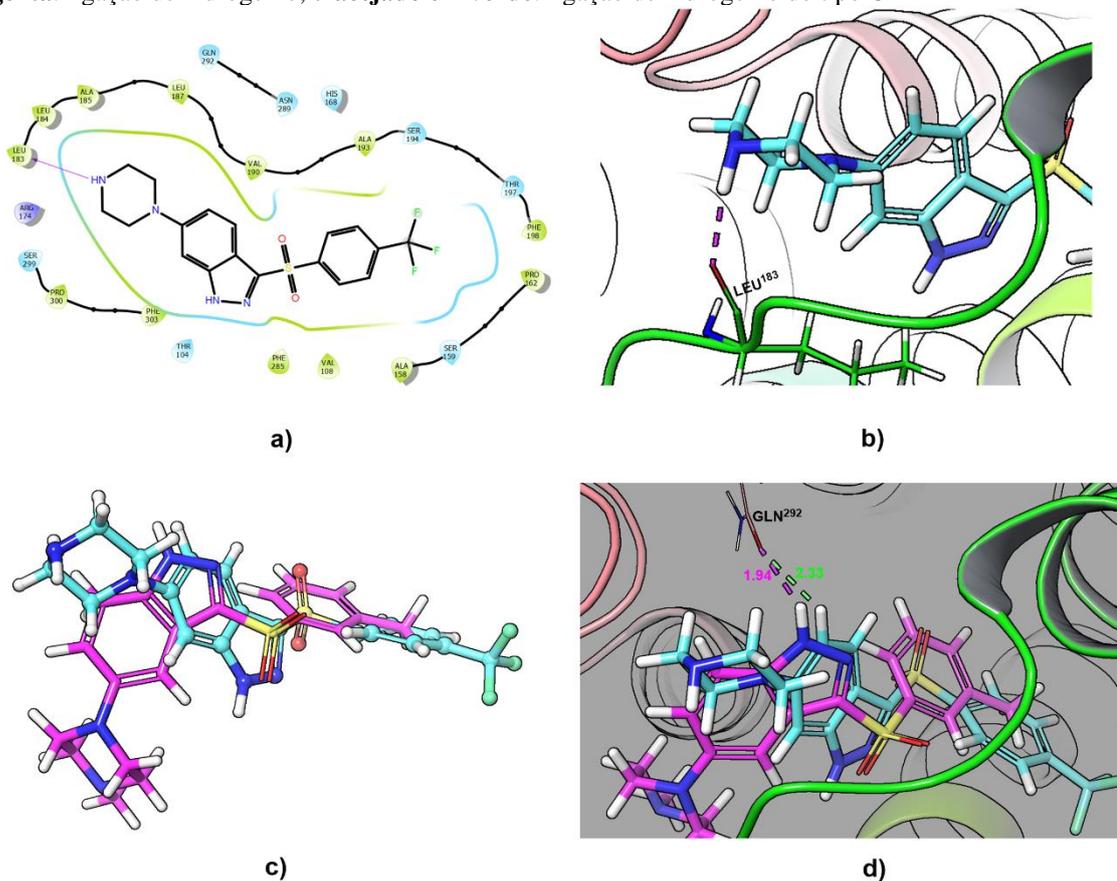
Fonte: autoria própria

Estudos prévios sobre a ativação de um GPCR<sup>167</sup> mostram que o movimento do segmento extracelular (região fora da membrana) da **TM6** em direção a **TM3** é essencial na ativação do

receptor 5-HT<sub>6</sub>. Observa-se a partir da Figura 57 um possível mecanismo molecular de bloqueio do receptor 5-HT<sub>6</sub> em que o anel benzimidazol (destacado em laranja) do composto **C15** interage com os aminoácidos **GLN**<sup>292</sup> e **ASN**<sup>289</sup> e esta interação impede que a **TM6** (Hélice em laranja) se desloque em direção a **TM3** (Hélice azul ciano), inativando o receptor. O resíduo **LEU**<sup>187</sup> presente no *loop* **ECL2** também interage com o anel benzimidazol e contribui para bloquear o movimento tanto da **TM6** quanto do **ECL2**, impulsionando também inativação do receptor. O grupo tolueno (destacado na cor verde-limão) está entre as **TMs 3 e 5** (Hélice representada na cor verde-limão) e interage com os resíduos **SER**<sup>194</sup> e **PHE**<sup>198</sup>, ambos localizados na **TM5**, e com o **VAL**<sup>108</sup> presente na **TM3** resultando na imobilização dessas duas hélices, cujos movimentos contribuem para ativar o receptor 5-HT<sub>6</sub>. Já o grupo metilpiperazina (destacado na cor azul ciano) está inserido entre as **TMs 3 e 6**, interagindo com o resíduo **ASP**<sup>107</sup> da **TM3** impedindo-a de movimentar-se em direção a **TM7**, a qual também tem um papel importante na ativação do receptor.

A simulação de *Docking* molecular para o composto com menor efeito inibitório (**C25**) revelou um modo de ligação diferente do observado para o **C15**. Conforme visto (Figura 58a-b), o grupo piperazina do **C25** interage por ligação de hidrogênio com o resíduo **LEU**<sup>183</sup> pertencente ao **ECL2**, alça que conecta as **TMs 4 e 5**. A grande semelhança entre os modos de ligação desses dois compostos (**C15** e **C25**) é a interação com o resíduo **GLN**<sup>292</sup> (Figura 58c-d). Contudo, enquanto o **C15** interage por ligação de hidrogênio convencional, o **C25** interage por ligação de hidrogênio do tipo CH... $\pi$ , também conhecida como interação alquila-arila, sendo esta última um tipo de interação mais fraca quando comparada ao primeiro tipo. O resíduo **GLN**<sup>292</sup> é importante, pois pertence a **TM6** e como mencionado contribui para a inativação do receptor. Todavia, a interação mais fraca entre esse resíduo e o **C25** provavelmente é a razão do baixo potencial inibitório desse composto ( $pK_i = 6,68$  - Tabela 2).

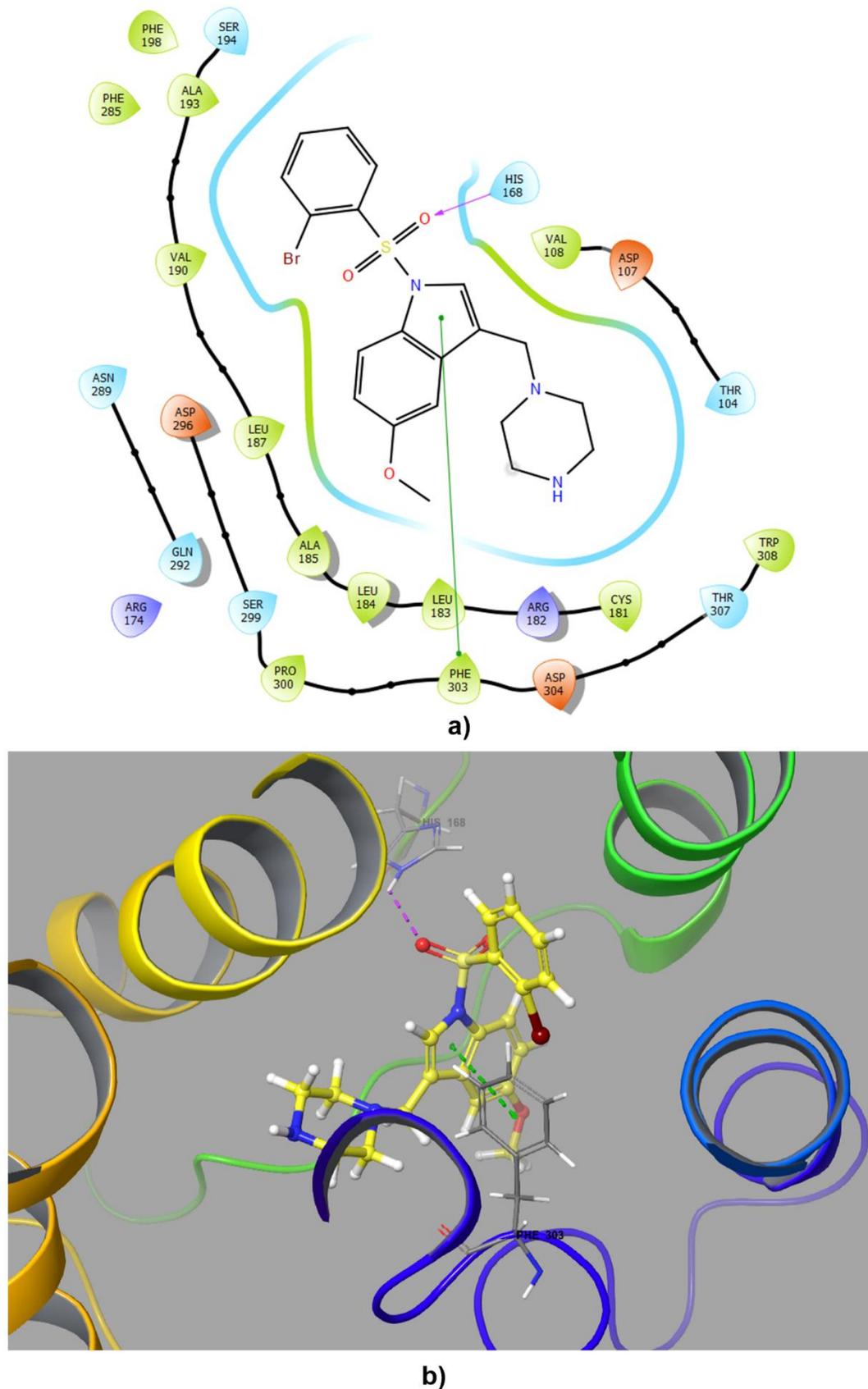
**Figura 58** - Comparação entre os modos de ligação dos compostos **C15** (estrutura em magenta) e **C25** (estrutura em azul) - **a)** representação bidimensional dos principais resíduos de aminoácidos envolvidos na interação com o **C25**; **b)** ligação de hidrogênio entre o resíduo **LEU<sup>183</sup>** e o **C25**; **c)** comparação entre as poses de ligação entre os compostos **C15** e **C25**; **d)** Diferença no modo de interação entre os compostos e o resíduo **GLN<sup>292</sup>**. **Tracejado em magenta:** ligação de hidrogênio; **tracejado em verde:** ligação de hidrogênio do tipo **CH... $\pi$** .



Fonte: autoria própria

Na seleção de compostos do conjunto 1 (Tabela 2) para realizar simulações de *Docking* molecular, escolheu-se o composto mais potente e o menos ativo, o mesmo critério de seleção foi aplicado ao conjunto 2 (Tabela 3), resultando na seleção do composto com maior e menor efeito antagonista (**C70** e **C67**, respectivamente), cujos tipos de interação podem ser visualizados nas Figuras 59 e 60.

**Figura 59** - Principais interações entre o composto **70** e o sítio ativo do receptor 5-HT<sub>6</sub> – a) diagrama bidimensional, b) representação tridimensional



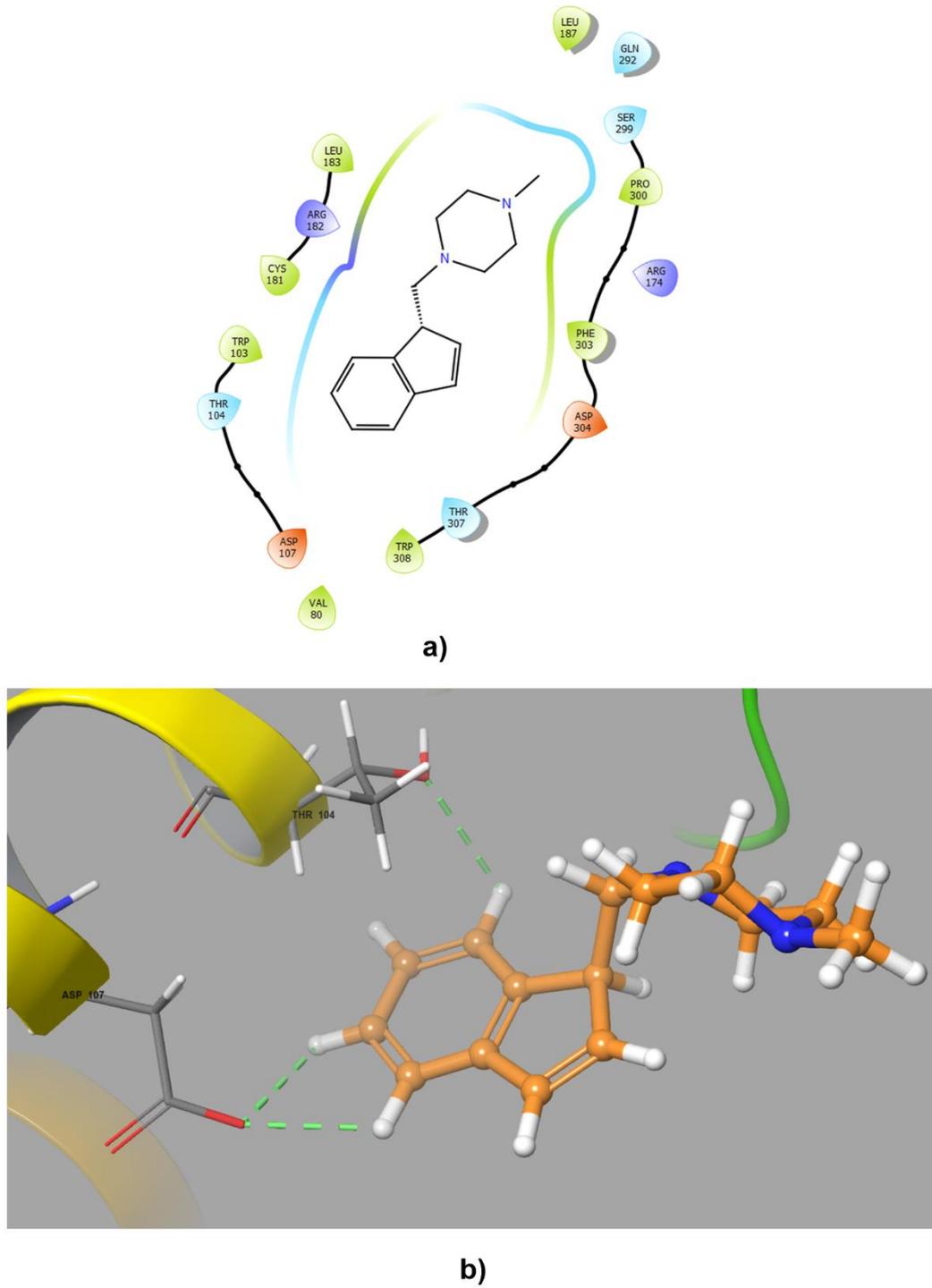
Fonte: autoria própria

Observa-se a partir da Figura 59a que o substituinte metóxi *indole* está embebido por um pacote hidrofóbico (LEU<sup>184</sup>, ALA<sup>185</sup>, LEU<sup>187</sup>, CYS<sup>181</sup>, PRO<sup>300</sup>, PHE<sup>303</sup>), cujos resíduos de aminoácidos fazem parte das **TM5** a **TM7** e conforme discutido, o bloqueio dessas TMs é essencial para a inativação do receptor 5-HT<sub>6</sub>. Entre os tipos de interações que esses resíduos participam, a interação mais evidente é a do tipo empilhamento  $\pi$  (também conhecida como *Pi-Pi stacking*) que existe entre o resíduo PHE<sup>303</sup> e o grupo metóxi *indole* (Figura 59b), interação essa que vem chamando muito atenção, devido a sua importância no processo de desenvolvimento de fármacos.<sup>168</sup>

As interações que ocorrem entre o grupo piperazina e os resíduos de aminoácidos THR<sup>104</sup>, ASP<sup>107</sup> e VAL<sup>108</sup> e THR<sup>307</sup> valem apenas serem pontuadas, pois esses resíduos fazem parte das **TMs 3** e **7**, cujos movimentos culminaria na ativação do receptor. No entanto, essas interações impedem o deslocamento da **TM3** em direção a **TM7** e esse bloqueio resulta na inativação do receptor 5-HT<sub>6</sub>. Vale mencionar também a interação entre o resíduo HIS<sup>168</sup> presente no *loop* extracelular 2 (ECL2) e o grupo sulfonil, agindo como uma espécie de amarra das **TM4** e **TM5**, e, portanto, impedindo o deslocamento dessas hélices na região extracelular.

Avaliou-se também as interações entre o receptor 5-HT<sub>6</sub> e o composto menos ativo (**C67**), e o principal tipo de interação observada foi CH... $\pi$ , ligação de hidrogênio mais fraca do que a convencional, conforme apontado ao longo dessa Seção 5.4. Os resíduos de aminoácidos envolvidos na interação com o grupo indeno (Figura 60a-b) são THR<sup>104</sup> e ASP<sup>107</sup> (dupla interação CH... $\pi$  com o anel indeno), ambos integrantes da **TM3**, hélice com papel destacado no processo de ativação do receptor. No entanto, esse tipo de interação por ser mais fraca do que as observadas para o composto mais ativo (**C70**) justifica o menor efeito antagonista do composto **C67**.

**Figura 60** - Principais tipos de interações entre os resíduos de aminoácidos e o composto **C67**- **a)** representação bidimensional, **b)** representação tridimensional



Fonte: autoria própria

## 5.5 Planejamento de novos compostos

A partir das análises dos modelos, bem como das informações extraídas das simulações de *Docking* molecular, novos compostos foram planejados. Primeiro, desenhou-se novas moléculas (Tabela 14) sob a ótica dos modelos QSAR-2D/3D gerados com o conjunto 1 de compostos (Tabela 2), já as análises realizadas para o modelo QSAR-ANN direcionaram o planejamento de outros possíveis antagonistas do receptor 5-HT<sub>6</sub> (Tabela 15). Acredita-se que o emprego de diferentes metodologias na geração dos modelos QSARs pode aumentar as chances de se obter protótipos com maiores probabilidades de serem candidatos a fármacos.

**Tabela 14** -compostos planejados a partir das modificações estruturais nas moléculas C15, C47 e C25  
**Novos compostos obtidos a partir das alterações na molécula C15**

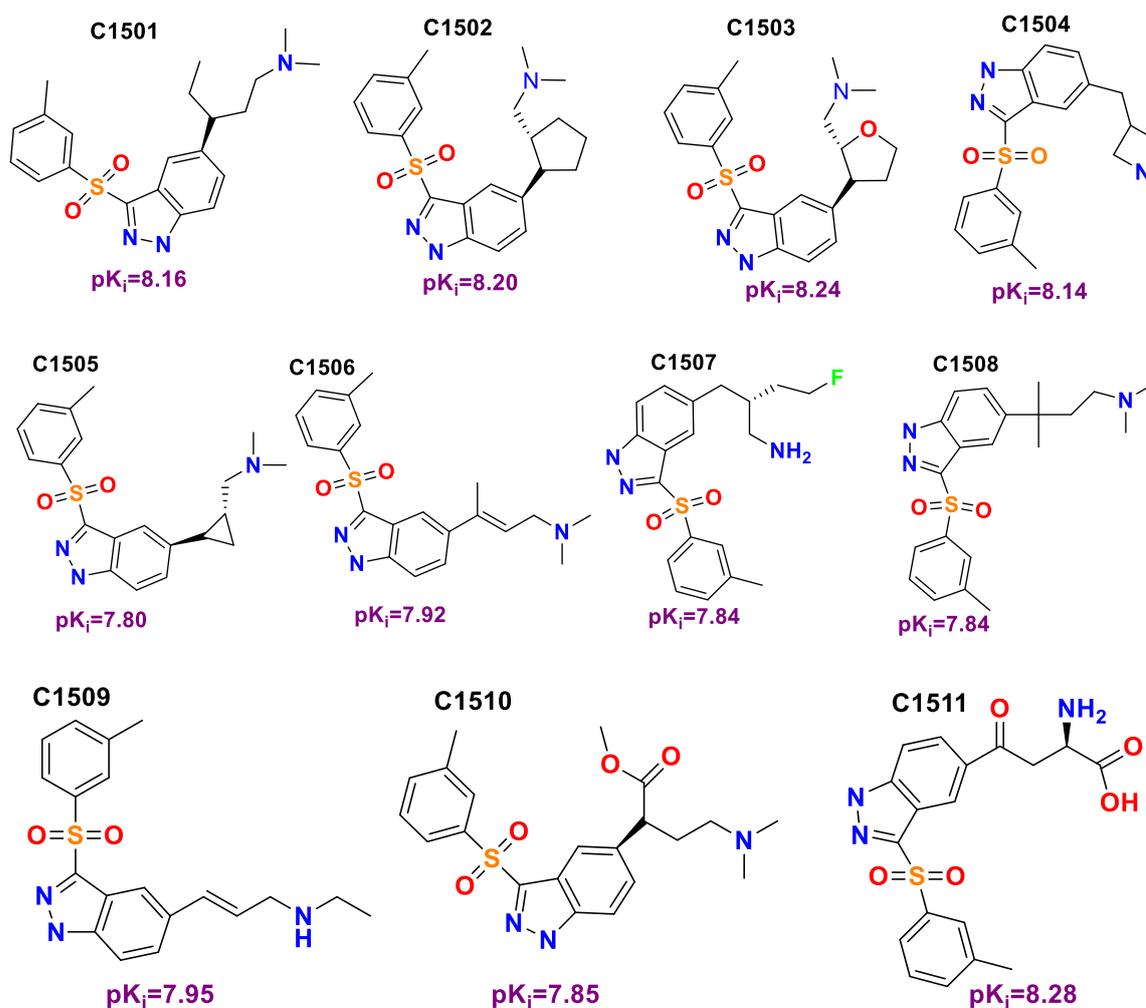
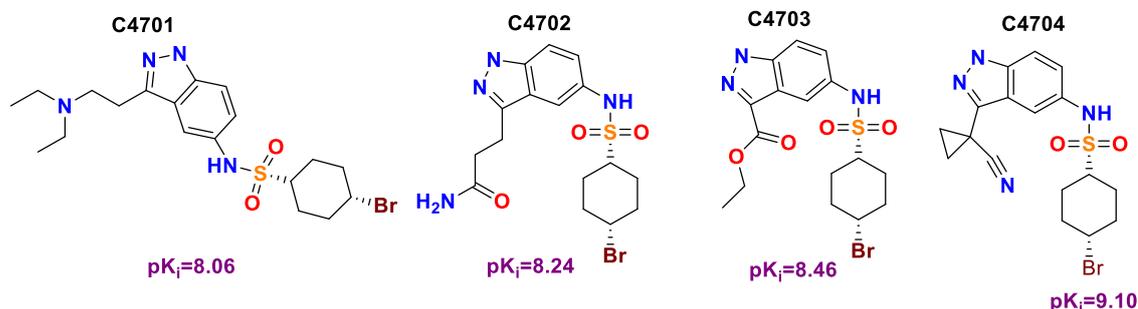
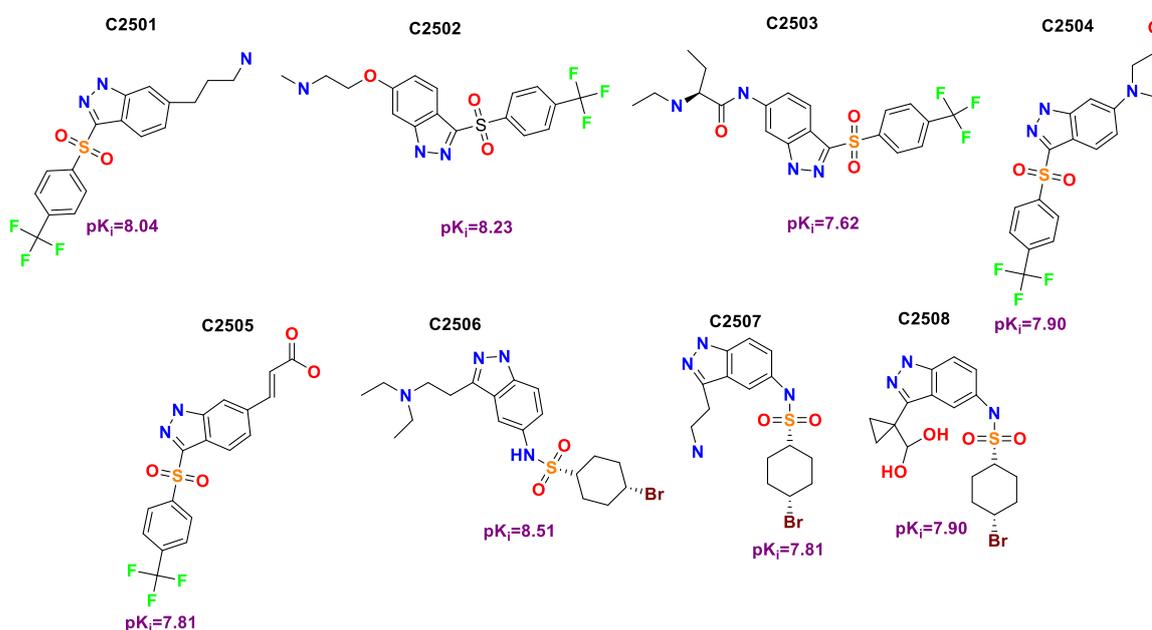


Tabela 14 - Continuação

## Novos compostos obtidos a partir das alterações na molécula C47



## Novos compostos obtidos a partir das alterações na molécula C25

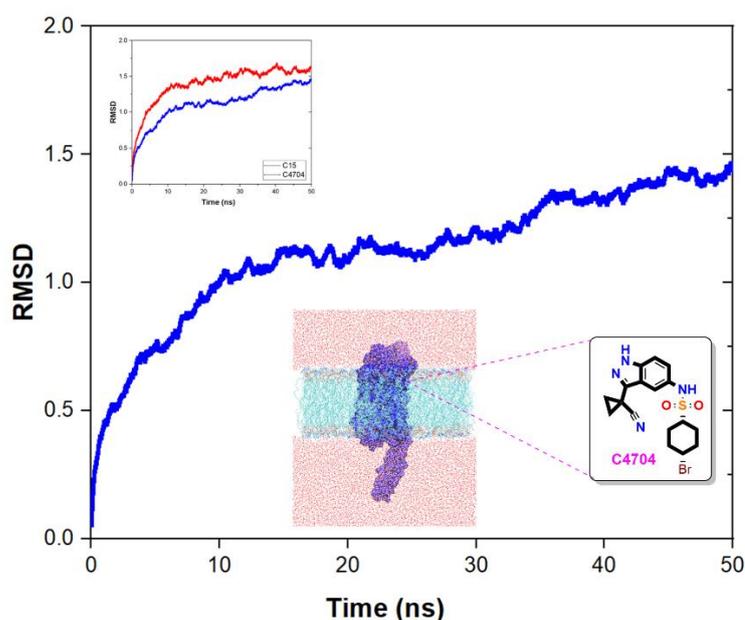


Fonte: autoria própria

Escolheu-se a partir do conjunto 1 (Tabela 2) os compostos **C15**, **C25**, **C47** (maior, moderado e menor efeito antagonista, respectivamente), para realizar modificações estruturais e posteriormente, os valores de  $pK_i$  dos novos compostos foram preditos usando os modelos QSARs aqui gerados. Como resultado, os compostos oriundos das modificações no **C15** não apresentaram valores de  $pK_i$  maior do que o composto de partida (**C15**), contudo esses novos compostos exibiram excelentes valores de atividade biológica. Já os novos compostos provenientes do **C47** exibiram significativos valores de  $pK_i$  quando comparados ao composto de partida. Em relação às moléculas desenhadas a partir do **C25**, todas elas exibiram valores de  $pK_i$  maiores do que o composto original. Vale a pena destacar que algumas mudanças estruturais realizadas a partir do composto **C47** gestaram uma molécula (**C4704** – Tabela 14) com valor de  $pK_i$  maior do que os valores atribuídos aos compostos do conjunto 1 (Tabela 2).

Selecionou-se a partir desse novo conjunto de compostos aquele com maior  $pK_i$  (**C4704**) predito para proceder para as simulações de *Dinâmica* molecular. Os resultados ratificam seu efeito antagonista realçado quando comparado ao composto mais ativo do conjunto 1 (**C15**), conforme observa-se na Figura 61.

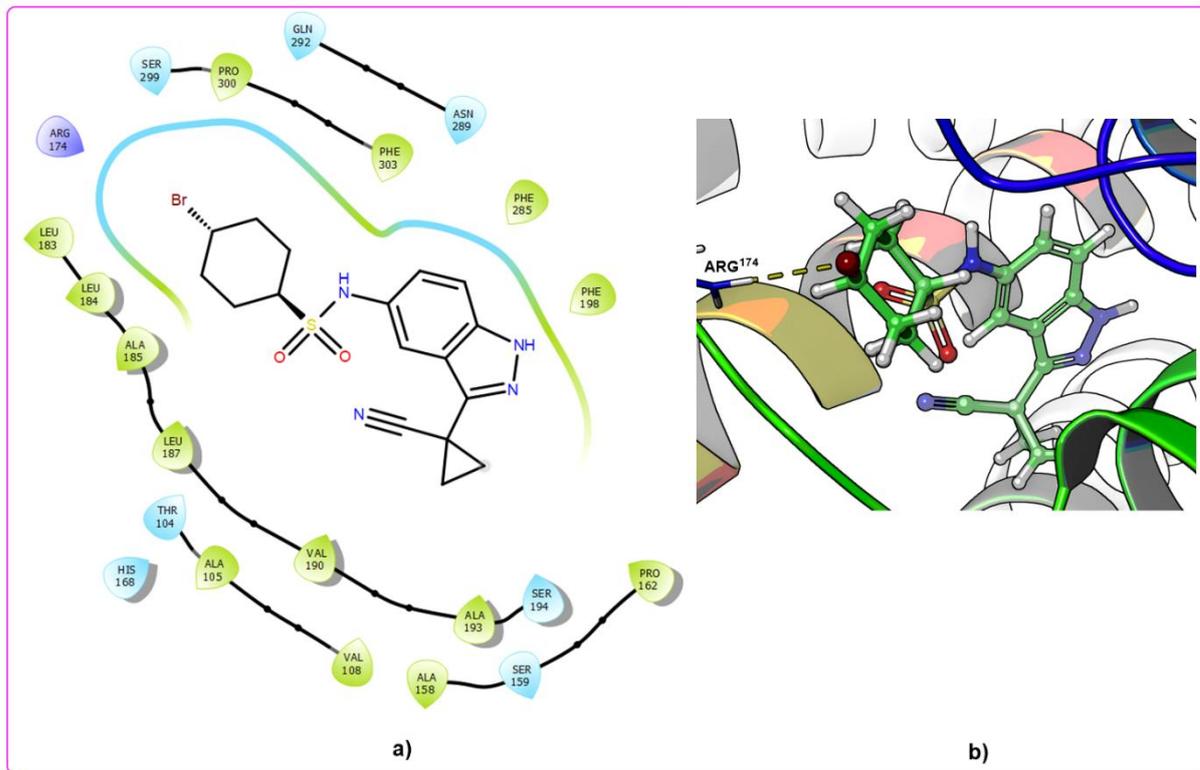
**Figura 61-** Simulação de DM para o complexo receptor-ligante (**C4704**) embebido em seu ambiente fisiológico. No canto esquerdo do gráfico tem-se uma comparação entre os RMSDs para as simulações com **C15** (gráfico em vermelho e composto **C4704** (representação em azul)



Fonte: autoria própria

Uma característica particular desse composto (**C4704**) quando comparado ao **C15** é um tipo especial de interação que ele faz com o receptor 5-HT<sub>6</sub>. As informações extraídas a partir das análises dos descritores apontavam para a importância de átomos de halogênios nas estruturas dos compostos, no entanto os resultados das simulações de *Docking* molecular não suscitaram essa importância, destacavam outras características que também contribuem para o efeito inibitório dos compostos. Levando-se em consideração as análises dos descritores, propôs-se moléculas com átomos de halogênios em suas estruturas e usou-se os modelos QSARs para prever a atividade biológica dessas novas moléculas. Posteriormente, escolheu-se o composto mais ativo para a simulação de DM e em seguida procedeu-se para as simulações de *Docking* molecular, usando o mesmo protocolo descrito na Seção 4.4. Os resultados mostraram que o composto **C4704**, diferentemente do composto **C15**, interage por ligação de halogênio (Figura 62), um tipo de interação bastante explorada no campo de planejamento de fármacos.<sup>169</sup>

**Figura 62-** Simulação de *Docking* molecular entre o receptor 5-HT<sub>6</sub> e composto **C4704**. **a)** representação bidimensional dos resíduos de aminoácidos que interagem com o **C4704**; **b)** principal tipo de interação entre o composto **C4704** e o resíduo ARG<sup>174</sup>



Fonte: autoria própria

Compostos halogenados vêm sendo amplamente empregados pela indústria farmacêutica para aumentar a permeabilidade na membrana ao mesmo tempo que diminui a degradação metabólica. Além do mais, as interações conhecidas como ligação de halogênio contribuem para aumentar a eficácia de vários fármacos.<sup>170</sup> De fato, vários dos novos compostos aqui planejados possuem átomos de halogênios em suas estruturas (**C4701**, **C4702**, **C4703**, **C4704**, **C2502**, **C2506**), cujos valores de atividade biológica preditos pelos modelos QSARs aqui gerados se mostraram promissores. Esses novos compostos quando submetidos a análises experimentais e em se confirmando esses ótimos efeitos antagonista podem ser excelentes candidatos anti-amnésicos.

A partir das análises do modelo ANN-QSAR aqui gerado, apostou-se em outra estratégia para se planejar novos antagonistas do receptor 5-HT<sub>6</sub>. Modificações nos grupos AR, HYD e PI foram conduzidas baseadas nas análises dos descritores selecionados para a construção do referido modelo em conjunto com as informações extraídas das interações receptor-ligante. Posteriormente, verificou-se se esses novos compostos apresentaram melhoras nos valores de pK<sub>i</sub> (Tabela 15):

Na região AR, o grupo *indole* foi substituído pelo grupo *indazole* (alteração realizada para todos os compostos planejados);

Várias modificações foram realizadas na região PI;

Diversas modificações foram conduzidas na região HYD, enquanto outras regiões permaneceram inalteradas;

Em vários momentos, modificações foram conduzidas nas regiões PI e HYD simultaneamente.

Tabela 15 - Compostos propostos a partir das análises do modelo ANN-QSAR

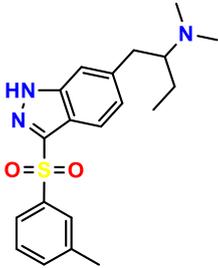
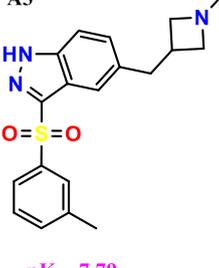
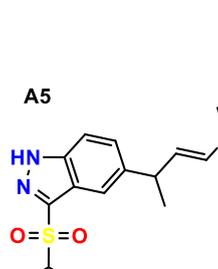
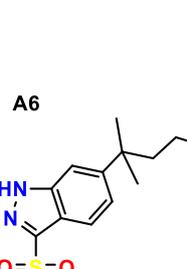
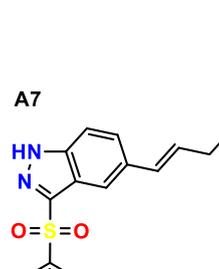
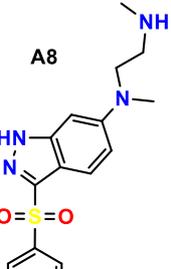
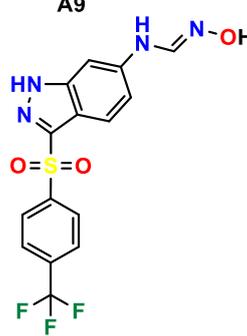
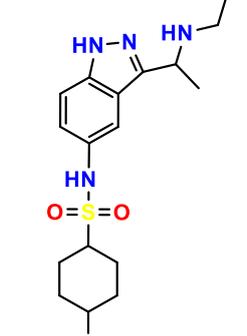
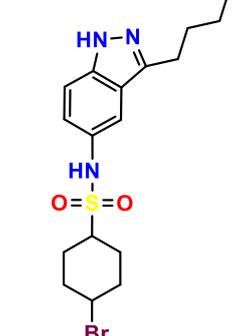
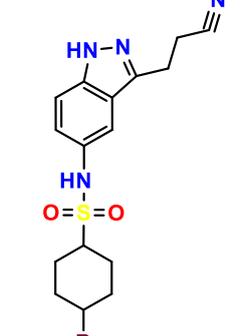
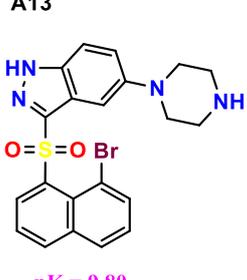
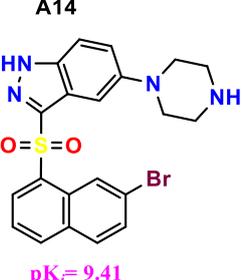
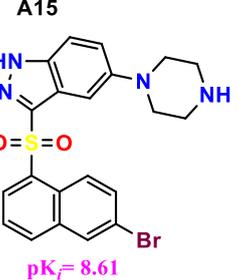
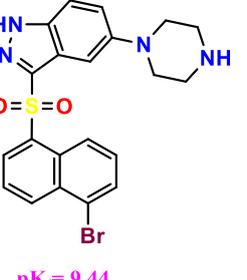
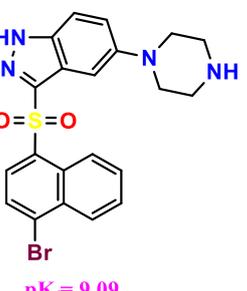
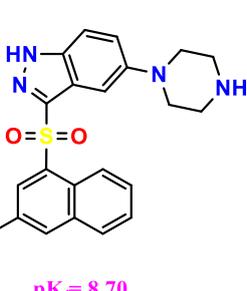
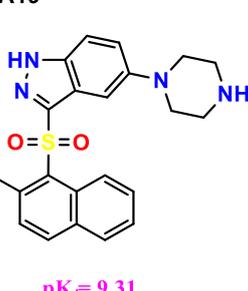
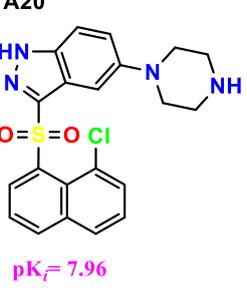
<p>A1</p>  <p><math>pK_i = 8.06</math></p>	<p>A2</p>  <p><math>pK_i = 7.61</math></p>	<p>A3</p>  <p><math>pK_i = 7.79</math></p>	<p>A4</p>  <p><math>pK_i = 8.03</math></p>
<p>A5</p>  <p><math>pK_i = 7.60</math></p>	<p>A6</p>  <p><math>pK_i = 6.56</math></p>	<p>A7</p>  <p><math>pK_i = 6.92</math></p>	<p>A8</p>  <p><math>pK_i = 5.54</math></p>

Tabela 15 – Continuação

<p><b>A9</b></p>  <p><math>pK_i = 6.53</math></p>	<p><b>A10</b></p>  <p><math>pK_i = 9.00</math></p>	<p><b>A11</b></p>  <p><math>pK_i = 6.78</math></p>	<p><b>A12</b></p>  <p><math>pK_i = 7.77</math></p>
<p><b>A13</b></p>  <p><math>pK_i = 9.80</math></p>	<p><b>A14</b></p>  <p><math>pK_i = 9.41</math></p>	<p><b>A15</b></p>  <p><math>pK_i = 8.61</math></p>	<p><b>A16</b></p>  <p><math>pK_i = 9.44</math></p>
<p><b>A17</b></p>  <p><math>pK_i = 9.09</math></p>	<p><b>A18</b></p>  <p><math>pK_i = 8.70</math></p>	<p><b>A19</b></p>  <p><math>pK_i = 9.31</math></p>	<p><b>A20</b></p>  <p><math>pK_i = 7.96</math></p>

Para os compostos **A1**, **A2**, **A3**, **A4**, **A5**, **A6** e **A7** (Tabela 15), as regiões AR e HYD permaneceram inalteradas, enquanto modificações graduais foram efetuadas na região PI. Entre os compostos **A1** a **A7**, destaca-se os ótimos valores de atividade biológica calculados para **A1** ( $pK_i = 8.06$ ), **A3** ( $pK_i = 7.79$ ) e **A4** ( $pK_i = 8.03$ ) preditos pelo modelo ANN. Como discutido (Seção 5.2.3), grupos volumosos na posição PI não são bem aceitos, razão pela qual os compostos com essas características não obtiveram valores de  $pK_i$  razoáveis (**A8**, **A9** e **A11** – Tabela 15). Insistiu-se em adicionar átomos de halogênios (Br, Cl e F) na região HYD, o que culminou em moléculas com promissores efeitos antagonistas, como por exemplo **A10**, **A13**, **A14**, **A15**, **A16**, **A17**, **A18**, e **A19**. Contudo, outras não apresentaram melhoras em seus valores de  $pK_i$ , como **A8** e **A9**, ambas com trifluor na região HYD, confirmando o efeito negativo de grupos volumosos nessa região sobre a resposta biológica, característica também observada

conforme os descritores iam sendo analisados. Como resultado geral, observou-se que a presença de átomos de halogênios em especial Bromo ligados ao anel aromático na posição *para* (região HYD) contribuem significativamente para aumentar o efeito antagonista dos compostos, logo, esses compostos podem ser ótimos pontos de partida para estudos experimentais posteriores.

### 5.6 Avaliação do perfil ADME/TOX dos novos compostos

Finalmente, analisou-se o perfil ADME/TOX dos novos compostos aqui propostos. Essas propriedades dão um indicativo do quão parecido são esses compostos (em termos de propriedades farmacocinéticas) com fármacos existentes no mercado. A Tabela 16 apresenta o perfil ADME/TOX desses compostos e os resultados mostram que eles apresentam boas chances, caso o efeito antagonista seja comprovado quando submetidos à abordagem experimental, de se tornarem fármacos.

**Tabela 16-** Estimativa do perfil ADEME/TOX dos novos compostos

<b>Propriedades ADME/Tox dos compostos planejados a partir do modelo QSAR-2D/3D</b>										
Composto	MW	HB <sub>doador</sub>	HB <sub>aceptor</sub>	QPlogPo/w	QPlogS	QPPCaco	QPlogBB	PHOA	PSA	R05
C1501	385.523	0.000	6.000	3.561	-4.217	199.946	-0.676	88.977	69.423	0
C1502	397.534	0.000	6.000	3.376	-4.129	148.282	-0.537	85.568	71.824	0
C1503	385.480	0.000	7.700	1.995	-2.348	154.026	-0.369	77.782	77.922	0
C1504	355.454	0.000	4.000	3.858	-4.857	489.171	-0.912	100.000	74.030	0
C1505	369.481	0.000	6.000	2.998	-3.718	179.785	-0.465	84.855	70.830	0
C1506	369.481	0.000	6.000	3.200	-4.163	188.698	-0.590	86.416	70.670	0
C1507	375.460	2.000	5.000	3.026	-4.256	62.688	-1.078	76.828	88.990	0
C1508	385.523	0.000	6.000	3.219	-3.612	134.405	-0.688	83.889	73.660	0
C1509	355.454	1.000	5.500	2.919	-3.753	118.136	-0.751	81.126	79.581	0
C1510	415.506	0.000	8.000	2.546	-2.989	89.921	-0.970	76.825	101.367	0
C1511	387.409	2.000	8.000	-0.704	-3.383	1.022	-2.353	22.990	161.987	0
C2501	383.388	2.000	5.000	2.802	-4.210	45.820	-0.961	73.080	93.050	0
C2502	399.387	1.000	6.250	3.015	-4.397	120.263	-0.585	81.828	87.870	0
C2503	454.466	2.000	8.000	3.043	-5.031	73.328	-0.963	78.145	108.236	0
C2504	399.387	1.000	6.700	2.897	-4.659	251.341	-1.125	86.867	90.025	0
C2505	396.340	1.000	6.000	3.061	-5.108	12.617	-1.879	64.572	119.233	0
C2506	457.428	2.000	7.500	2.978	-3.531	285.626	-0.320	88.335	76.586	0
C2507	401.320	4.000	6.500	1.540	-2.792	83.133	-0.730	70.320	96.226	0
C2508	444.342	4.000	8.900	1.576	-4.072	234.265	-1.310	78.588	110.525	0
C4701	457.428	2.000	7.500	3.142	-4.355	224.979	-0.507	87.442	75.782	0
C4702	429.331	4.000	8.000	0.868	-2.866	45.125	-1.743	61.639	125.429	0
C4703	430.316	1.000	6.500	2.888	-5.212	278.254	-1.238	87.606	106.051	0
C4704	423.327	2.000	7.000	2.447	-5.928	246.001	-1.198	84.069	97.160	0
<b>Propriedades ADME/Tox dos compostos planejados a partir do modelo QSAR-ANN</b>										
A1	371.496	0.00	6.000	2.879	-3.396	137.849	-0.715	82.095	72.673	0
A2	383.507	1.00	5.500	3.457	-4.939	115.551	-0.719	84.108	79.380	0
A3	355.454	0.000	4.000	4.234	-5.837	674.102	-0.919	100.000	71.112	0
A4	355.454	1.00	5.500	2.969	-4.324	134.002	-0.640	82.402	79.490	0
A5	369.481	0.000	6.000	3.441	-4.761	587.204	-0.988	96.651	70.923	0
A6	385.523	0.000	6.000	3.511	-4.164	187.242	-0.608	88.172	70.444	0
A7	341.427	1.00	5.500	2.761	-4.106	118.084	-0.785	80.199	79.209	0
A8	424.440	1.00	6.500	3.287	-5.098	131.158	-0.338	84.093	80.935	0
A9	384.332	2.000	7.700	1.392	-4.292	32.691	-2.028	62.204	134.109	0
A10	429.374	3.000	7.000	2.400	-3.750	151.099	-0.525	80.004	86.854	0
A11	428.386	2.000	5.500	3.862	-5.782	579.288	-1.052	100.000	79.617	0

Tabela 16 - Continuação

Composto	MW	HB <sub>doador</sub>	HB <sub>aceptor</sub>	QPlogP <sub>o/w</sub>	QPlogS	QPPCaco	QPlogBB	PHOA	PSA	R05
A12	411.316	2.000	7.000	2.247	-5.345	184.071	-1.379	80.638	100.401	0
A13	471.371	1.000	6.500	3.051	-4.419	114.591	-0.440	81.667	82.595	0
A14	471.371	1.000	6.500	3.342	-4.876	131.733	-0.308	84.453	79.699	0
A15	471.371	1.000	6.500	3.327	-4.908	123.735	-0.335	83.878	79.806	0
A16	471.371	1.000	6.500	3.255	-4.761	123.703	-0.342	83.451	79.763	0
A17	471.371	1.000	6.500	3.262	-4.770	124.519	-0.340	83.549	79.787	0
A18	471.371	1.000	6.500	3.318	-4.898	121.730	-0.345	83.699	79.905	0
A19	471.371	1.000	6.500	2.986	-4.189	116.321	-0.380	81.398	79.566	0
A20	426.920	1.000	6.500	3.010	-4.363	114.824	-0.443	81.440	82.853	0

**Legenda:**

**MW**- Peso molecular (130.0 a 725.0 Daltons);

**HB<sub>doador</sub>** – Número de doadores de hidrogênio recomendado (0.0 a 6.0);

**HB<sub>aceptor</sub>** – Número de aceptores de hidrogênio recomendado (2.0 a 20.0);

**QPlogP<sub>o/w</sub>** – Coeficiente de partição álcool/água predito (-2.0 a 6.5);

**QPlogS** – Solubilidade aquosa predita, log S (-6.5 a 0.5);

**QPPCaco** – Predição da permeabilidade nas células Caco-2 em nm/sec (<25-pobre, > 500-ótima);

**QPlogBB** – Predição do coeficiente cérebro/sangue (-3,0 a 1,2);

**PHOA** – Predição da absorção oral humana em uma escala de 0 a 100% (> 80% é ótima, <25% é ruim);

**PSA** - área de superfície de van der Waals de átomos polares, N e O, além dos átomos da carboníla (70 -200.0);

**R05** – Regra dos cinco de Lipinski (Mol\_MW<500, QPlogP<sub>o/w</sub><5, HB<sub>doador</sub>≤5 e HB<sub>aceptor</sub>≤ 10).

Um estudo realizado por Lipinski e colaboradores<sup>61</sup> usando 2 mil fármacos revelou um padrão farmacocinético - para que um *lead* composto tenha maior probabilidade de se tornar um fármaco, ele deve apresentar as seguintes características: **i) MW <500, ii) Qplog/w < 5, iii) HB<sub>doador</sub> ≤ 5, iv) HB<sub>aceptor</sub> ≤ 10**. Compostos que atendam a essas características, denominada de regra dos cinco de Lipinski (R05, por serem múltiplos de 5), são considerados compostos drogáveis. Sendo assim, levando em consideração essa regra, a Tabela 16 mostra que nenhum dos compostos propostos a violou, e, portanto, em se confirmando experimentalmente o efeito antagonista, eles possuem boas chances de serem candidatos a fármacos.

Ao longo do processo de desenvolvimento de um fármaco, um dos parâmetros analisados é a permeabilidade intestinal, a qual é estudada usando as células Caco-2 que quando cultivadas apresentam características da mucosa intestinal e por isso, estudos com essas células pontuarão o grau de absorção de um composto no organismo.<sup>171</sup> Sendo assim, o emprego da abordagem *in silico* pode auxiliar no processo de triagem de compostos, uma vez que há uma faixa adequada de valores de coeficiente de permeabilidade intestinal (coeficiente menor que 25 nm/s é considerado pobre, ao passo que valores maiores do que 500 nm/s são considerados ótimos em absorção). A Tabela 16 mostra que apenas o composto **C1511** apresenta valor abaixo do requerido, e, portanto, não seria uma boa opção para proceder para os estudos experimentais, apesar do ótimo valor de atividade biológica.

Os compostos aqui propostos (Tabelas 14 e 15) foram planejados para atuarem no sistema nervoso central, mas como discutido ao longo desse documento, fármacos com esse direcionamento devem contornar alguns percalços, entre eles têm-se a barreira

hematoencefálica que impede a invasão de corpos estranhos no SNC. Pensando nisso, desenvolveu-se modelos capazes de prever se um determinado composto está numa faixa de valores adequados (-3 a 1,2 – valores abaixo de -3 possuem pouca chance de atravessar a BB, ao passo que quanto mais próximo de 1,2 maior a chance de o composto atravessar a barreira hematoencefálica). Em estando e se confirmando os ótimos valores de atividade biológica por meio de procedimentos experimentais, esses compostos podem ser considerados ótimos candidatos a fármacos. A Tabela 16 mostra que todos os compostos estão nessa faixa de valores permitida e, portanto, são ótimos pontos de partida para futuras abordagens experimentais.

Como última análise, mensurou-se a acessibilidade sintética dos novos compostos (Tabela 17), em que é medido o grau de dificuldade para sintetizá-los (score de 1 a 10, em que 1 é extremamente fácil e 10 significa que é a síntese é muito difícil). Os resultados dessa avaliação indicam que os novos compostos não apresentam grau de dificuldade sintética elevado, e, portanto, uma vez que todas as outras análises aqui realizadas apontam esses compostos como drogáveis, acredita-se que eles são excelentes pontos de partida para futuros estudos experimentais.

**Tabela 17-** Predição da acessibilidade sintética para os novos compostos

Compostos oriundos do modelo QSAR 2D/3D				Compostos oriundos do modelo QSAR-ANN			
Composto	Score	Composto	score	Composto	Score	Composto	score
<b>C1501</b>	3,61	<b>C1502</b>	3,95	<b>A1</b>	5,20	<b>A11</b>	4,17
<b>C1503</b>	3,99	<b>C1504</b>	2,88	<b>A2</b>	3,87	<b>A12</b>	3,92
<b>C1505</b>	3,71	<b>C1506</b>	3,27	<b>A3</b>	2,88	<b>A13</b>	3,27
<b>C1507</b>	3,49	<b>C1508</b>	3,18	<b>A4</b>	3,60	<b>A14</b>	3,18
<b>C1509</b>	3,23	<b>C1510</b>	3,70	<b>A5</b>	3,78	<b>A15</b>	3,15
<b>1511</b>	3,29	<b>C4701</b>	4,29	<b>A6</b>	3,22	<b>A16</b>	3,18
<b>C4702</b>	3,91	<b>C4703</b>	3,93	<b>A7</b>	3,13	<b>A17</b>	5,16
<b>C4704</b>	3,96	<b>C2501</b>	2,86	<b>A8</b>	3,00	<b>A18</b>	3,27
<b>C2502</b>	2,86	<b>C2503</b>	3,52	<b>A9</b>	2,78	<b>A19</b>	3,33
<b>C2504</b>	2,84	<b>C2505</b>	2,86	<b>A10</b>	4,21	<b>A20</b>	3,24
<b>C2506</b>	4,29	<b>C2507</b>	3,89				
<b>C2508</b>	4,09						

## 6 CONCLUSÃO

Modelos matemáticos são amplamente empregados por diversos ramos da ciência para estimar vários fenômenos. Nesta pesquisa de doutorado, usou-se diferentes ferramentas computacionais para construir modelos QSARs e assim analisar a relação entre a estrutura química de vários antagonistas do receptor 5HT<sub>6</sub> e seus valores de atividade biológica. Posteriormente, a partir das informações extraídas desses modelos, bem como das análises receptor-ligante via simulação de *Docking* molecular, novos potenciais antagonistas desse receptor foram planejados.

Para todos os modelos QSARs gerados (QSAR 2D/3D e ANN), vários procedimentos de validação foram aplicados para garantir a robustez e a capacidade preditiva de todos eles, como por exemplo, procedimento de validação cruzada, teste de randomização, e como prova indubitável sobre a performance preditiva de ambos, empregou-se procedimento de validação externa com a utilização de conjuntos de compostos externos.

Informações acerca da interação receptor-ligante também contribuíram para o planejamento dos novos compostos. Para isso, um modelo tridimensional do receptor 5-HT<sub>6</sub> foi construído e posteriormente validado via procedimento de *Dinâmica* molecular. Ressalta-se ainda que a confiabilidade nos resultados da simulação de *Docking* molecular só foi possível após simulação do receptor 5-HT<sub>6</sub> em seu ambiente fisiológico utilizando DM.

A aplicação dos métodos PLS e CoMFA na construção de modelos QSARs resultaram no planejamento de compostos cujos valores preditos de pK<sub>i</sub> são maiores do que os valores dos compostos usados na construção do modelo, valendo destacar o composto **C4704** com ótimo valor de pK<sub>i</sub>. A fim de aumentar as chances de se obter antagonistas do receptor 5-HT<sub>6</sub> com maior probabilidade de serem candidatos a fármacos, empregou-se a técnica de redes neurais artificiais na construção de um outro modelo QSAR. Como resultado obteve-se compostos (**A10**, **A13**, **A14**, **A16**, **A17**, **A18** e **A19**) com valores preditos de pK<sub>i</sub> excelentes (pK<sub>i</sub> > **9,00**).

Conclui-se a partir das análises dos descritores, em conjunto com as informações extraídas das interações receptor-ligante, que a presença de átomos de halogênios (principalmente Br) ligados a anéis aromáticos na posição *para* (região HYD) contribui para aumentar significativamente o efeito antagonista dos compostos, já grupos demasiadamente volumosos na região PI não resultam em compostos com efeito inibitório apreciáveis.

Por último, as ótimas propriedades ADME/Tox, bem como a avaliação da acessibilidade sintética de todos os compostos aqui planejados, obtidas via abordagem *in silico*, os qualificam

para futuros estudos experimentais, abordagem essa que poderá certificar os efeitos inibitórios desses compostos, bem como suas características drogáveis.

## REFERÊNCIAS

- 1 WILSON, R. S; SEGAWA, E; BOYLE, P. A; ANAGNOS, S. E; HIZEL, L. P; BENNETT, D.A. The natural history of cognitive decline in Alzheimer's disease. **Psychology and Aging**, Chicago, v. 27, n. 4, p. 1008–1017, 2012.
- 2 BARKER, W. W; LUIS, C.A; KASHUBA, A; LUIS, M; HARWOOD, D. G; LOEWENSTEIN, D; WATERS, C; JIMISON, P; SHEPHERD, E; SEVUSH, S; GRAFF-RADFORD, N; NEWLAND, D; TODD, M; MILLER, B; GOLD, M; HEILMAN, K; DOTY, L; GOODMAN, I; ROBINSON, B; PEARL, G; DICKSON, D; DUARA, R. Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank. **Alzheimerer Disease & Associated Disorder**, Philadelphia, v.16, n. 4, p. 203-212, 2002.
- 3 KATZMAN, R. The Prevalence and Malignancy of Alzheimer Disease. A major killer. **Alzheimer's & Dementia**, Hoboken, v. 33, n.4, p. 217–218, 1979.
- 4 GAUGLER, J.; JAMES, B.; JOHNSON, T.; SCHOLZ, K.; WEUVE, J. 2016 Alzheimer's disease facts and figures. **Alzheimer's & Dementia**, Hoboken, v.12, n. 4, p. 459-509, 2016.
- 5 STRASSNIG, M.; GANGULI, M. About a peculiar disease of the cerebral cortex: Alzheimer's original case revisited. **Psychiatry**, Pittsburgh, v. 2, n. 9, p. 30-33, 2005.
- 6 HIPPIUS; H.; NEUNDÖRFER, G. The discovery of Alzheimer's disease. **Dialogues in Clinical Neuroscience**, Neuilly sur Seine, v. 5, n. 1, p. 101- 108, 2003.
- 7 WALDEMAR, G.; BURN, A. **Alzheimer's Disease**. 2 ed. Oxford: Oxford University Press, 2017. 144 p. ISBN 978-019-87980-30.
- 8 GORDON, B. A.; BLAZEY, T. M.; SU, Y.; HARI-RAJ, A.; DINCER, A.; FLORES, S.; CHRISTENSEN, J.; MCDADE, E.; WANG, G.; XIONG, C.; CAIRNS, N. J.; HASSENSTAB, J.; MARCUS, D. S.; FAGAN, A. M.; JACK, C. R. JR.; HORNBECK, R .C.; PAUMIER, K. L.; ANCES, B. M.; BERMAN, S. B.; BRICKMAN, A. M.; CASH, D. M.; CHHATWAL, J. P.; CORREIA, S.; FÖRSTER, S.; FOX, N. C.; GRAFF-RADFORD, N. R.; LA FOUGÈRE, C.; LEVIN, J.; MASTERS, C. L.; ROSSOR, M. N.; SALLOWAY, S.; SAYKIN, A. J.; SCHOFIELD, P. R.; THOMPSON, P. M.; WEINER, M. M.; HOLTZMAN, D. M.; RAICHLE, M. E.; MORRIS, J. C.; BATEMAN, R. J.; BENZINGER, T. L. S. Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. **The Lancet Neurology**, London, v. 17, n. 3, p. 241-250, 2018.
- 9 VILLEMAGNE, V. L.; BURNHAM, S.; BOURGEAT, P.; BROWN, B.; ELLIS, K. A.; SALVADO, O.; SZOEKE, C.; MACAULAY, S. L.; MARTINS, R.; MARUFF, P.; AMES, D.; ROWE, C. C.; MASTERS, C. L. Amyloid  $\beta$  deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. **Lancet Neurology**, London, v. 12, n. 4, p. 357-367, 2013.

- 10 REIMAN, E. M.; QUIROZ, Y. T.; FLEISHER, A. S.; CHEN, K.; VELEZ-PARDO, C.; JIMENEZ-DEL-RIO, M.; FAGAN, A. M.; SHAH, A. R.; ALVAREZ, S.; ARBELAEZ, A.; GIRALDO, M.; ACOSTA-BAENA, N.; SPERLING, R. A.; DICKERSON, B.; STERN, C. E.; TIRADO, V.; MUNOZ, C.; REIMAN, R. A.; HUENTELMAN, M. J.; ALEXANDER, G. E.; LANGBAUM, J. B.; KOSIK, K. S.; TARIOT, P. N.; LOPERA, F. Brain imaging and fluid biomarker analysis in young adults at genetic risk for autosomal dominant Alzheimer's disease in the presenilin 1 E280A kindred: A case-control study. **Lancet Neurology**, London, v. 11, n. 12, p. 1048-1056, 2012.
- 11 BRAAK, H.; THAL, D. R.; GHEBREMEDHIN, E.; DEL TREDICI, K. Stages of the pathologic process in alzheimer disease: age categories from 1 to 100 years. **Journal of Neuropathology & Experimental Neurology**, Cary, v. 70, n. 11, p. 960-969, 2011.
- 12 ALZHEIMER'S DISEASE INTERNATIONAL. **World Alzheimer Report 2019: attitudes to dementia**. London, 2019. Disponível em: <https://www.alzint.org/u/WorldAlzheimerReport2019.pdf>. Acesso em: 15 jan. 2020.
- 13 GBD 2016 DEMENTIA COLLABORATORS. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. **Lancet Neurology**, London, v. 18, n.1, p. 88-106. jan. 2019.
- 14 VALADARES, C. **A cada três segundos, um idoso é diagnosticado com algum tipo de demência no mundo**. Brasília: Blog da Saúde. Ministério da Saúde, 2017. Disponível em: <http://www.blog.saude.gov.br/index.php/geral/51880-a-cada-tres-segundos-um-idoso-e-diagnosticado-com-algum-tipo-de-demencia-no-brasil>. Acesso em: 9 mar. 2020.
- 15 MAL do Alzheimer: conscientização como caminho para inclusão. **G1**, Fortaleza, 2018. Disponível em: <https://g1.globo.com/ce/ceara/especial-publicitario/sistema-fecomercio/radar-do-comercio/noticia/2018/09/13/mal-do-alzheimer-conscientizacao-como-caminho-para-inclusao.ghtml> Acesso em: 9 mar. 2020.
- 16 BARBARINO, P.; LYNCH, C.; BLISS, A.; DABAS, L. **From plan to impact II - The urgent need for action**. London: Alzheimer's Disease International, 2019. Disponível em: <https://www.alz.co.uk/adi/pdf/from-plan-to-impact-2019.pdf>. Acesso em: 9 mar. 2020.
- 17 MEMORY T. H. **Brain neurons and synapses**. [S. l.]: The Human Memory, 2020. Disponível em: <https://human-memory.net/brain-neurons-synapses/>. Acesso em: 10 de set. 2020
- 18 TEPLow, D. B.; YANG, M.; ROYCHAUDHURI, R.; PANG, E.; HUYNH, T. P.; CHEN, M. S. B. S. **Alzheimer's disease: targets for new clinical diagnostic and therapeutic strategies- frontiers in neuroscience**. Boca Raton: CRC Press, 2012. p. 298. ISBN 978-1-4398-2708-6.
- 19 PAUL, E. Tau imaging in preclinical Alzheimer's disease. *In*: PERNECZKY, R. **Biomarkers for preclinical Alzheimer's disease**. New York: Humana Press, 2018. p. 272. ISBN 978-1-4939-9255-3.
- 20 DE STROOPER, B.; KARRAN, E. The cellular phase of Alzheimer's disease. **Cell**, Cambridge, v. 164, n. 4, p. 603–615, 2016.

- 21 BENILOVA, I.; KARRAN, E.; DE STROOPER, B. The toxic A $\beta$  oligomer and Alzheimer's disease: an emperor in need of clothes. **Nature Neuroscience**, New York, v. 15, n. 3, p. 349–357, 2012.
- 22 HAASS, C.; SCHLOSSMACHER, M. G.; HUNG, A. Y.; VIGO-PELFREY, C.; MELLON, A.; OSTASZEWSKI, B. L.; LIEBERBURG, I.; KOO, E. H.; SCHENK, D.; TEPLow, D. B.; SELKOE, D. J. Amyloid  $\beta$ -peptide is produced by cultured cells during normal metabolism. **Nature**, London, v. 359, p. 322–325, 1992.
- 23 MATTSON, M. P. Cellular actions of beta-amyloid precursor protein and its soluble and fibrillogenic derivatives. **Physiological Reviews**, Rockville, v. 77, n. 4, p. 1081–132, 1997.
- 24 SOREGHAN, B.; KOSMOSKI, J.; GLABE, C. Surfactant properties of Alzheimer's A beta peptides and the mechanism of amyloid aggregation. **The Journal of Biological Chemistry**, Bethesda, v. 269, n. 46, p. 28551–28554, 1994.
- 25 DEIBEL, M.; EHMANN, W.; MARKESBERY, W. Copper, iron, and zinc imbalances in severely degenerated brain regions in Alzheimer's disease: possible relation to oxidative stress. **Journal of the Neurological Sciences**, Amsterdam, v. 143, n. 1–2, p. 137–142, 1996.
- 26 HANE, F.; LEONENKO, Z. Effect of metals on kinetic pathways of amyloid- $\beta$  aggregation. **Biomolecules**, Basel, v. 4, n. 1, p. 101–116, 2014.
- 27 BUSH, A.; PETTINGELL, W.; MULTHAUP, G.; D PARADIS, M.; VONSATTEL, J.; GUSELLA, J.; BEYREUTHER, K.; MASTERS, C. L.; TANZI, R. E. Rapid induction of Alzheimer A beta amyloid formation by zinc. **Science**, Washington, v. 265, n. 5177, p. 1464–1467, 1994.
- 28 BUSH, A. I. The metal theory of Alzheimer's disease. **Journal of Alzheimer's Disease**, Amsterdam, n. 33, p. 277–281, 2013.
- 29 FINEFROCK, A. E.; BUSH, A. I.; DORAISWAMY, P. M. Current status of metals as therapeutic targets in Alzheimer's disease. **Journal of the American Geriatrics Society**, Malden, v. 51, n. 8, p. 1143–1148, 2003.
- 30 BUSH, A. I. The metallobiology of Alzheimer's disease. **Trends in Neurosciences**, Oxford, v. 26, n. 4, p. 207–214, 2003.
- 31 PARDRIDGE, W. M. Alzheimer's disease drug development and the problem of the blood-brain barrier. **Alzheimer's & Dementia**, Hoboken, v. 5, n. 5, p. 427–432, 2009.
- 32 DEDEOGLU, A.; CORMIER, K.; PAYTON, S.; TSEITLIN, K. A.; KREMSKY, J. N.; LAI, L.; LI, X.; MOIR, R. D.; TANZI, R. E.; BUSH, A. I.; KOWALL, N. W.; ROGERS, J. T.; HUANG, X. Preliminary studies of a novel bifunctional metal chelator targeting Alzheimer's amyloidogenesis. **Experimental Gerontology**, Philadelphia, v. 39, n. 11–12, p. 1641–1649, 2004.
- 33 KEBERLE, H. The biochemistry of desferrioxamine and its relation to iron metabolism. **Annals of the New York Academy of Sciences**, New York, v. 119, n. 2, p. 758–768, 1964.

- 34 CUAJUNGCO, M. P.; FAGÉT, K. Y.; HUANG, X.; TANZI, R. E.; BUSH, A. I. Metal chelation as a potential therapy for Alzheimer's disease. **Annals of the New York Academy of Sciences**, New York, v. 920, n. 1, p. 292–304, 2006.
- 35 RICHARDSON, D. R.; PONKA, P. Development of iron chelators to treat iron overload disease and their use as experimental tools to probe intracellular iron metabolism. **American Journal of Hematology**, Hoboken, v. 58, n. 4, p. 299–305, 1998.
- 36 OLIVIERI, N. F.; BRITTENHAM, G. M. Iron-chelating therapy and the treatment of thalassemia. *Erratu*. **Blood**, Washington, v. 89, n.7, p. 2621, 1997. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/9028304/>. Acesso em: 37 abr. 2020.
- 37 KONTOGHIORGHES, G. J. New concepts of iron and aluminium chelation therapy with oral L1 (deferiprone) and other chelators. A review. **The Analyst**, Cambridge, v. 120, n. 3, p. 845-851, 1995.
- 38 BARTUS, R.; DEAN, R.; BEER, B.; LIPPA, A. The cholinergic hypothesis of geriatric memory dysfunction. **Science**, United States v. 217, n. 4558, p. 408–414, 1982.
- 39 RAWLINS, M. D. Cutting the cost of drug development? **Nature Reviews Drug Discovery**, London, v. 3, n. 4, p. 360–364, 2004.
- 40 BENHAMÚ, B.; MARTÍN-FONTECHA, M.; VÁZQUEZ-VILLA, H.; PARDO, L.; LÓPEZ-RODRÍGUEZ, M. L. Serotonin 5-HT<sub>6</sub> receptor antagonists for the treatment of cognitive deficiency in Alzheimer's disease. **Journal of Medicinal Chemistry**, Washington, v. 57, n. 17, p. 7160-7181, 2014.
- 41 KUDO, T.; DAVIS, K. L.; GONZALEZ, R. B.; WILKINSON, D. G. **Practical pharmacology for Alzheimer's disease**. New York: Springer International Publishing, 2016. 141 p. ISBN 978-3-319-26206-2.
- 42 COYLE, J.; PRICE, D.; DELONG, M. Alzheimer's disease: a disorder of cortical cholinergic innervation. **Science**, Washington, v. 219, n. 4589, p. 1184-1890, 1983.
- 43 KÁSA, P.; RAKONCZAY, Z.; GULYA, K. The cholinergic system in Alzheimer's disease. **Progress in Neurobiology**, Oxford, v. 52, n. 6, p. 511-535, 1997.
- 44 FERRIS, S.; IHL, R.; ROBERT, P.; WINBLAD, B.; GATZ, G.; TENNIGKEIT, F.; GAUTHIER, S. Treatment effects of memantine on language in moderate to severe Alzheimer's disease patients. **Alzheimer's & Dementia**, Hoboken, v. 5, n. 5, p. 369-374, 2009.
- 45 SUCHER, N. J.; AWOBULUYI, M.; CHOI, Y. B.; LIPTON, S. A. NMDA receptors: from genes to channels. **Trends in Pharmacological Sciences**, Oxford, v. 17, n. 10, p. 348-355, 1996.
- 46 DINGLELINE, R.; BORGES, K.; BOWIE, D.; TRAYNELIS, S. F. The glutamate receptor ion channels. **Pharmacological Reviews**, Rockville, v. 51, n. 1, p. 7–61, 1999.

- 47 GREENAMYRE, J. T.; YOUNG, A. B. Excitatory amino acids and Alzheimer's disease. **Neurobiology of Aging**, Philadelphia, v. 10, n. 5, p. 563-602, 1989.
- 48 ZENARO, E.; PIACENTINO, G.; CONSTANTIN, G. The blood-brain barrier in Alzheimer's disease. **Neurobiology of Disease**, Maryland Heights, v. 107, p. 41–56, 2017.
- 49 ZLOKOVIC, B. V. The blood-brain barrier in health and chronic neurodegenerative disorders. **Neuron**, Cambridge, v. 57, n. 2, p. 178-201, 2008.
- 50 BALLABH, P.; BRAUN, A.; NEDERGAARD, M. The blood–brain barrier: an overview: Structure regulation, and clinical implications. **Neurobiology of Disease**, Maryland Heights, v. 16, n.1, p.1-13, 2004.
- 51 WALLEZ, Y; HUBER, P. Endothelial adherens and tight junctions in vascular homeostasis, inflammation and angiogenesis. **Biochimica et Biophysica Acta: biomembranes**, Amsterdam, v. 1778, n. 3, p. 794–809, 2008.
- 52 TIETZ, S; ENGELHARDT, B. Brain barriers: Crosstalk between complex tight junctions and adherens junctions. **The Journal Cell Biology**, New York, v. 209, n. 4, p. 493-506, 2015.
- 53 MONTAGNE, A; BARNES, S. R; SWEENEY, M. D; HALLIDAY, M. R; SAGARE, A. P; ZHAO, Z; TOGA, A. W; JACOBS, R. E; LIU, C. Y; AMEZCUA, L; HARRINGTON, M. G; CHUI, H. C; LAW, M; ZLOKOVIC, B. V. Blood-brain barrier breakdown in the aging human hippocampus. **Neuron**, Cambridge, v. 85, n. 2, p. 296-302, 2015.
- 54 ROHER, A. E; KUO, Y. M; ESH, C; KNEBEL, C; WEISS, N; KALBACK, W; LUEHRS, D. C; CHILDRESS, J. L; BEACH, T. G; WELLER, R. O; KOKJOHN, K. T. Cortical and leptomeningeal cerebrovascular amyloid and white matter pathology in Alzheimer's disease. **Molecular Medicine**, London, v. 9, n. 3-4, p. 112–22, 2003.
- 55 ERICKSON, M. A; BANKS, W. A. Blood–brain barrier dysfunction as a cause and consequence of Alzheimer's disease. **Journal of Cerebral Blood Flow and Metabolism**, London, v. 33, n. 10, p. 1500–13, 2013.
- 56 FORMAN, M. S; LAL, D; ZHANG, B; DABIR, D. V; SWANSON, E; LEE, V. M-Y; TROJANOWSKI, J.Q. Transgenic mouse model of Tau pathology in astrocytes leading to nervous system degeneration. **The Journal of Neuroscience**, Washington, v. 25, n. 14, p. 3539–3550, 2005.
- 57 BLAIR, L. J; FRAUEN, H. D; ZHANG, B; NORDHUES, B. A; BIJAN, S; LIN, Y. C; ZAMUDIO, F; HERNANDEZ, L. D; SABBAGH, J. J; SELENICA, M. L; DICKEY, C. A. Tau depletion prevents progressive blood-brain barrier damage in a mouse model of tauopathy. **Acta Neuropathologica Communications**, London, v. 3, n. 1, p. 8, 2015.
- 58 KOVAC, A; ZILKOVA, M; DELI, M. A; ZILKA, N; NOVAK, M. Human truncated Tau is using a different mechanism from Amyloid- $\beta$  to damage the blood-brain barrier. **Journal of Alzheimer's Disease**, Amsterdam, v. 18, n. 4, p. 897–906, 2009.

- 59 VIDAL, R; CALERO, M; PICCARDO, P; FARLOW, M. R; UNVERZAGT, F. W; MÉNDEZ, E; JIMÉNEZ-HUETE, A.; BEAVIS, R; GALLO, G; GOMEZ-TORTOSA, E; GHISO, J; HYMAN, B. T; FRANGIONE, B; GHETTI, B. Senile dementia associated with amyloid  $\beta$  protein angiopathy and tau perivascular pathology but not neuritic plaques in patients homozygous for the APOE- $\epsilon$ 4 allele. **Acta Neuropathologica**, Heidelberg, v. 100, n. 1, p.1–12, 2000.
- 60 SIMPKINS, J. W. Commentary on “Alzheimer’s disease drug development and the problem of the blood-brain barrier. **Alzheimer’s & Dementia**, Hoboken, v. 5, n. 5, p. 433-434, 2009.
- 61 LIPINSKI, C. A.; LOMBARDO, F.; DOMINY, B. W.; DOMINY, B.W.; FEENEY, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, Amsterdam, v. 46, n. 1-3, p. 3-26, 1997.
- 62 COHEN, B. E.; BANGHAM, A. D. Diffusion of small non-electrolytes across liposome membranes. **Nature**, London, v. 236, n. 5343, p.173–174, 1972.
- 63 FISCHER, H.; GOTTSCHLICH, R.; SEELIG, A. Blood-brain barrier permeation: molecular parameters governing passive diffusion. **The Journal of Membrane Biology**, New York, v. 165, n.3, p. 201–11, 1998.
- 64 RISHTON, G. M.; LABONTE, K.; WILLIAMS, A. J.; KASSAM, K.; KOLOVANOV, E. Computational approaches to the prediction of blood-brain barrier permeability: A comparative analysis of central nervous system drugs versus secretase inhibitors for Alzheimer’s disease. **Current Opinion Drug Discovery & Development**, London, v. 9, n.3, p. 303-313, 2006.
- 65 UPTON, N.; CHUANG, T. T.; HUNTER, A. J.; VIRLEY, D. J. 5-HT<sub>6</sub> receptor antagonists as novel cognitive enhancing agents for Alzheimer’s disease. **Neurotherapeutics**, New York, v. 5, n.3, p. 458–69, 2008.
- 66 KARILA, D; FRERET, T; BOUET, V; BOULOUARD, M; DALLEMAGNE, P; ROCHAIS, C. Therapeutic Potential of 5-HT<sub>6</sub> Receptor Agonists. **Journal of Medicinal Chemistry**, Washington, v. 58, n. 20, p. 7901-7912, 2015.
- 67 KOHEN, R.; METCALF, M. A.; KHAN, N.; DRUCK, T.; HUEBNER, K.; LACHOWICZ, J. E.; MELTZER, H. Y.; SIBLEY, D. R.; ROTH, B. L.; HAMBLIN, M. W. Cloning, characterization, and chromosomal localization of a human 5-HT<sub>6</sub> Serotonin Receptor. **Journal of Neurochemistry**, Chichester, v. 66, n. 1, p. 47–56, 2002.
- 68 MONSMA JUNIOR, F. J.; SHEN, Y.; WARD, R. P.; HAMBLIN, M. W.; SIBLEY, D. R. Cloning and expression of a novel serotonin receptor with high affinity for tricyclic psychotropic drugs. **Molecular Pharmacology**, Rockville, v. 43, n. 3, p. 320-307, 1993.
- 69 LEE, Y.; BASITH, S.; CHOI, S. Recent advances in structure-based drug design targeting class a g protein-coupled receptors utilizing crystal structures and computational simulations. **Journal of Medicinal Chemistry**, Washington, v. 61, n.1, p. 1–46, 2018.

- 70 OLDHAM, W. M.; HAMM, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. **Nature Reviews Molecular Cell Biology**, London, v. 9, n. 1, p. 60–71, 2008.
- 71 PLASSAT, J. L.; AMLAIKY, N.; HEN, R. Molecular cloning of a mammalian serotonin receptor that activates adenylate cyclase. **Molecular Pharmacology**, Rockville, v. 44, n. 2, p. 229-236, 1993.
- 72 WOOLLEY, M.; BENTLEY, J.; SLEIGHT, A.; MARSDEN, C.; FONE, K. C. A role for 5-HT<sub>6</sub> receptors in retention of spatial learning in the Morris water maze. **Neuropharmacology**, Oxford, v. 41, n. 2, p. 210–219, 2001.
- 73 BENTLEY, J. C.; BOURSON, A.; BOESS, F. G.; FONE, K. C.; MARSDEN, C. A.; PETIT, N. S. A. Investigation of stretching behaviour induced by the selective 5-HT<sub>6</sub> receptor antagonist, Ro 04-6790, in rats. **British Journal of Pharmacology**, Oxford, v. 126, n. 7, p. 1537–42, 1999.
- 74 DE LA FUENTE, T.; MARTÍN-FONTECHA, M.; SALLANDER, J.; BENHAMÚ, B.; CAMPILLO, M.; MEDINA, R. A.; PELLISSIER, L. P.; CLAEYSEN, S.; DUMUIS, A.; PARDO, L.; LÓPEZ-RODRÍGUEZ, M. L. Benzimidazole derivatives as new serotonin 5-HT<sub>6</sub> receptor antagonists. Molecular mechanisms of receptor inactivation. **Journal of Medicinal Chemistry**, Washington, v. 53, n. 3, p.1357–69, 2010.
- 75 BENTLEY, J. C.; BOURSON, A.; BOESS, F. G.; FONE, K. C.; MARSDEN, C. A.; PETIT, N.; SLEIGHT, A. J. Investigation of stretching behaviour induced by the selective 5-HT<sub>6</sub> receptor antagonist, Ro 04-6790, in rats. **British Journal of Pharmacology**, Oxford, v. 126, n.7, p. 1537-1542, 1999.
- 76 BOURSON, A.; BORRONI, E.; AUSTIN, R. H.; JR MONSMA, F. J.; SLEIGHT, A. Determination of the role of the 5-HT<sub>6</sub> receptor in the rat brain: a study using antisense oligonucleotides. **The Journal of Pharmacology and Experimental Therapeutics**, Rockville, v. 274, n. 1, p. 173-80, 1995.
- 77 SLEIGHT, A. J.; MONSMA, F. J.; BORRONI, E.; AUSTIN, R. H.; BOURSON, A. Effects of altered 5-HT<sub>6</sub> expression in the rat: functional studies using antisense oligonucleotides. **Behavioural Brain Research**, Amsterdam, v. 73, n. 1-2, p. 245-248, 1995.
- 78 WOOD, P. L.; MALTHER-SØRENSEN, D.; CHENEY, D. L.; COSTA, E. Increase of hippocampal acetylcholine turnover rate and the stretching-yawning syndrome elicited by alpha-MSH and ACTH. **Life Sciences**, Philadelphia, v. 22, n. 8, p. 673-678, 1978.
- 79 DAWSON, L. A.; NGUYEN, H. Q.; LI, P. *In vivo* effects of the 5-HT<sub>6</sub> antagonist SB-271046 on striatal and frontal cortex extracellular concentrations of noradrenaline, dopamine, 5-HT, glutamate and aspartate. **British Journal of Pharmacology**, Oxford, v. 130, n. 1, p. 23-26, 2000.
- 80 ROGERS, D. C.; HAGAN, J. J. 5-HT<sub>6</sub> receptor antagonists enhance retention of a water maze task in the rat. **Psychopharmacology**, Heidelberg, v. 158, n. 22, p. 114-119, 2001.

- 81 WOODS, S.; CLARKE, N.; LAYFIELD, R.; FONE, K. 5-HT<sub>6</sub> receptor agonists and antagonists enhance learning and memory in a conditioned emotion response paradigm by modulation of cholinergic and glutamatergic mechanisms. **British Journal of Pharmacology**, Oxford, v. 167, n. 2, p. 436-449, 2012.
- 82 LÓPEZ-RODRÍGUEZ, M. L.; BENHAMÚ, B.; DE LA FUENTE, T.; SANZ, A.; PARDO, L.; CAMPILLO, M. A Three-dimensional pharmacophore model for 5-hydroxytryptamine<sub>6</sub> (5-HT<sub>6</sub>) receptor antagonists. **Journal of Medicinal Chemistry**, Washington, v. 48, n. 13, p. 4216-4219, 2005.
- 83 LEITÃO, A.; MONTANARI, M. L. C.; MONTANARI, C. A. Desenvolvimento de fármacos. *In*: MONTANARI, C. A. (org.). **Química medicinal: métodos e fundamentos em planejamento de fármacos**, São Paulo: EDUSP. 2011. 732 p. ISBN 10: 8531412668.
- 84 ACHARYA, C.; COOP, A. E.; POLLI, J. D.; MACKERELL, A. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. **Current Compu-Aided-Drug Design**, Sharjah, v. 7, n. 1, p. 10-22, 2011.
- 85 DRWAL, M. N.; GRIFFITH, R. Combination of ligand- and structure-based methods in virtual screening. **Drug Discovery Today Technologies**, London, v. 10, n.3, p. 395-401, 2013.
- 86 GUIDO, R. V. C.; ANDRICOPULO, A. D. Modelagem molecular de fármacos. **Revista de Processos Químicos**, Goiânia, v. 2, n. 4, p. 24-36, 2008.
- 87 BERMAN, H. M; WESTBROOK, J; FENG, Z; GILLILAND, G; BHAT, T. N; WEISSIG, H; SHINDYALOV, I. N; BOURNE, P. E. The Protein Data Bank. **Nucleic Acids Research**, Oxford, v. 28, p. 235-242, 2000.
- 88 PATRICK, G. L. **An introduction to medicinal chemistry**. 5. ed. Oxford: Oxford University Press, 2013. 816 p. ISBN 978-019874991.
- 89 GUEDES, I. A.; MAGALHÃES, C. S.; DARDENNE, L. E. Atracamento molecular. *In*: VERLI, H. **Bioinformática: da biologia a flexibilidade moleculares**. São Paulo: SBBq, 2014. p. 282. ISBN 978-85-69288-00-8.
- 90 VERLI, H. Dinâmica molecular. *In*: VERLI, H. **Bioinformática: da biologia a flexibilidade moleculares**. São Paulo: SBBq, 2014. p. 173-187. ISBN 978-85-69288-00-8.
- 91 MARTNEZ, L.; BORIN, I. A.; SKAF, M. S. Fundamentos de simulação por dinâmica molecular. *In*: MORGON, N. H.; COUTINHO, K. **Métodos de química teórica e modelagem molecular**. São Paulo: Livraria da Física, 2007. 540 p. ISBN 978-858832587X.
- 92 DE VIVO, M.; MASETTI, M.; BOTTEGONI, G.; CAVALLI, A. Role of molecular dynamics and related methods in drug discovery. **Journal of Medicinal Chemistry**, Washington, v. 59, n. 9, p. 4035-4061, 2016.
- 93 JENSEN, F. **Introduction to computational chemistry**. 2. ed. Chichester: Wiley 2007. 599 p. ISBN-10: 0470011874.

- 94 CORNELL, W. D.; CIEPLAK, P.; BAYLY, C. I.; GOULD, I. R.; MERZ, K. M.; FERGUSON, D. M.; SPELLMEYER, D. C.; FOX, T.; CALDWELL, J. W.; KOLLMAN, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. **Journal of the American Chemical Society**, Washington, v. 117, n. 19, p. 5179-5197, 1995.
- 95 HANSEN, H. S.; HÜNENBERGER, P. H. A reoptimized GROMOS force field for hexopyranose-based carbohydrates accounting for the relative free energies of ring conformers, anomers, epimers, hydroxymethyl rotamers, and glycosidic linkage conformers. **Journal of Computational Chemistry**, Hoboken, v. 32, n. 6, p. 998-1032, 2011.
- 96 PIANA, S.; LINDORFF-LARSEN, K.; SHAW, D. E. How robust are protein folding simulations with respect to force field parameterization? **Biophysical Journal**, St. Louis, v. 100, n. 9, p. 47-49, 2011.
- 97 KONY, D.; DAMM, W.; STOLL, S.; VAN GUNSTEREN, W. F. An improved OPLS-AA force field for carbohydrates. **Journal Computational Chemistry**, Hoboken, v. 23, n. 15, p. 1416-1429, 2002.
- 98 HOLLINGSWORTH, S. A.; DROR, R. O. Molecular dynamics simulation for all. **Neuron**, Cambridge, v. 99, n. 6, p. 1129-1143, 2018.
- 99 KIM, S.; CHEN, J.; CHENG, T.; GINDULYTE, A.; HE, J.; HE, S.; LI, Q.; SHOEMAKER, B. A.; THIESSEN, P. A.; YU, B.; ZASLAVSKY, L.; ZHANG, J.; BOLTON, E. E. PubChem 2019 update: improved access to chemical data. **Nucleic Acids Research**, Oxford, v.47, p. 1102–1109, 2019.
- 100 GAULTON, A.; BELLIS, L. J.; BENTO, A. P.; CHAMBERS, J.; DAVIES, M.; HERSEY, A.; LIGHT, Y.; MCGLINCHEY, S.; MICHALOVICH, D.; AL-LAZIKANI, B.; OVERINGTON, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. **Nucleic Acids Research**, Oxford, v. 40, p. 1100–1107, 2012.
- 101 LU, P.; MAGWANGA, R. O.; KIRUNGU, J. N.; DONG, Q.; CAI, X.; ZHOU, Z.; WANG, X.; XU, Y.; PENG, R.; WANG, K.; LIU, F. GENOME-wide analysis of the cotton G-coupled receptor proteins (GPCR) and functional analysis of GTOM1, a novel cotton GPCR gene under drought and cold stress. **BMC Genomics**, London, v. 20, n. 651, p. 1-17, 2019.
- 102 HAO, M.; LI, Y.; LI, H.; ZHANG, S. Investigation of the structure requirement for 5-HT<sub>6</sub> binding affinity of arylsulfonyl derivatives: a computational study. **International Journal of Molecular Sciences**, Basel, v. 12, n. 8. p. 5011-5030. 2011.
- 103 RENTERÍA-GÓMEZ, A.; ISLAS-JÁCOME, A.; DÍAZ-CERVANTES, E.; VILLASEÑOR-GRANADOS, T.; ROBLES, J.; GÁMEZ-MONTAÑO, R. Synthesis of azepino[4,5-b]indol-4-ones via MCR/free radical cyclization and in vitro–in silico studies as 5-HT<sub>6</sub>R ligands. **Bioorganic & Medicinal Chemistry Letter**, Oxford, v. 26, n. 9, p. 2333-2338, 2016.

104 FABRITIUS, C. H.; PESONEN, U.; MESSINGER, J.; HORVATH, R.; SALO, H.; GAŁĘZOWSKI, M.; GALEK, M.; STEFAŃSKA, K.; SZEREMETA-SPISAK, J.; OLSZAK-PŁACHTA, M.; BUDA, A.; ADAMCZYK, J.; KRÓL, M.; PRUSIS, P.; SIEPRAWKA-LUPA, M.; MIKULSKI, M.; KUOKKANEN, K.; CHAPMAN, H.; OBUCHOWICZ, R.; KORJAMO, T.; JALAVA, N.; NOWAK, M. 1-Sulfonyl-6-Piperazinyl-7-Azaindoles as potent and pseudo-selective 5-HT<sub>6</sub> receptor antagonists. **Bioorganic & Medicinal Chemistry Letter**, Oxford, v. 26, n. 11, p. 2610-2615, 2016.

105 DUKAT, M; MOSIER, P. D; KOLANOS, R; ROTH, B. L; GLENNON, R. A. Binding of serotonin and N<sub>1</sub>-benzenesulfonyltryptamine-related analogs at human 5-HT<sub>6</sub> serotonin receptors: receptor modeling studies. **Journal of Medicinal Chemistry**, Washington, v. 51, n. 3, p. 603-611, 2008.

106 PENG, J.; XU, J. Raptorx: exploiting structure information for protein alignment by statistical inference. **Proteins: structure, functions and bioinformatics**, Hoboken, v. 79, n. 10, p. 161-171, 2011.

107 KÄLLBERG, M.; WANG, H.; WANG, S.; PENG, J.; WANG, Z.; LU, H.; XU, J. Template-based protein structure modeling using the RaptorX web server. **Nature Protocols**, London, v. 7, n. 8, p. 1511-1522, 2012.

108 CHERKASOV, A.; MURATOV, E. N.; FOURCHES, D.; VARNEK, A.; BASKIN, I. I.; CRONIN, M.; DEARDEN, J.; GRAMATICA, P.; MARTIN, Y. C.; TODESCHINI, R.; CONSONNI, V.; KUZ'MIN, V. E.; CRAMER, R.; BENIGNI, R.; YANG, C.; RATHMAN, J.; TERFLOTH, L.; GASTEIGER, J.; RICHARD, A.; TROPSHA, A. QSAR modeling: where have you been? Where are you going to? **Journal of Medicinal Chemistry**, Washington, v. 57, n. 12, p. 4977-5010, 2014.

109 MARTINS, J. P. A.; FERREIRA, M. M. C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. **Química Nova**, São Paulo, v. 36, n. 4, p. 554-560, 2013.

110 OECD. **Guidance document on the validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] models**. Paris, 2007. (OECD environment health and safety publications series on testing and assessment, 69). Disponível em: <https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm>. Acesso em: 12 ago. 2019.

111 KEISER, M. J.; SETOLA, V.; IRWIN, J. J.; LAGGNER, C.; ABBAS, A. I.; HUFSEIN, S. J.; JENSEN, N. H.; KUIJER, M. B.; MATOS, R. C.; TRAN, T. B.; WHALEY, R.; GLENNON, R. A.; HERT, J.; THOMAS, K. L.; EDWARDS, D. D.; SHOICHET, B. K.; ROTH, B. L. Predicting new molecular targets for known drugs. **Nature**, London, v. 462, n. 7270, p. 175-81, 2009.

112 LOUNKINE, E.; KEISER, M. J.; WHITEBREAD, S.; MIKHAILOV, D.; HAMON, J.; JENKINS, J. L.; LAVAN, P.; WEBER, E.; DOAK, A. K.; CÔTÉ, S.; SHOICHET, B. K.; URBAN, L. Large-scale prediction and testing of drug activity on side-effect targets. **Nature**, London, v. 486, n. 7403, p. 361-367, 2012.

113. BESNARD, J.; RUDA, G. F.; SETOLA, V.; ABECASSIS, K.; RODRIGUIZ, R. M.; HUANG, X.; NORVAL, S.; SASSANO, M. F.; SHIN, A. L.; WEBSTER, L. A.; SIMEONS, F. R. C.; STOJANOVSKI, L.; PRAT, A.; SEIDAH, N. G.; CONSTAM, D. B.; BICKERTON, R.; READ, K. D.; WETSEL, W. C.; GILBERT, I. H.; ROTH, B. L.; HOPKINS, A. L. Automated design of ligands to polypharmacological profiles. **Nature**, London, v. 492, n. 7428, p. 215–220, 2012.
114. HAJJO, R.; GRULKE, C. M.; GOLBRAIKH, A.; SETOLA, V.; HUANG, X.; ROTH, B. L.; TROPSHA, A. Development, validation, and use of Quantitative Structure–Activity Relationship models of 5-Hydroxytryptamine (2B) receptor ligands to identify novel receptor binders and putative valvulopathic compounds among common drugs. **Journal of Medicinal Chemistry**, Washington, v. 53, n. 21, p. 7573–7586, 2010.
115. COLE, D. C.; LENNOX, W. J.; LOMBARDI, S.; ELLINGBOE, J. W.; BERNOTAS, R. C.; TAWA, G. J.; MAZANDARANI, H.; SMITH, D. L.; ZHANG, G.; COUPET, J.; SCHECHTER, L. E. Discovery of 5-Arylsulfonamido-3- (pyrrolidin-2-ylmethyl)-1 *H* -indole derivatives as potent, selective 5-HT<sub>6</sub> receptor agonists and antagonists. **Journal of Medicinal Chemistry**, Washington, v. 48, n. 2, p. 353-356, 2005.
116. KUZ'MIN, V. E.; ARTEMENKO, A. G.; LOZITSKY, V. P.; MURATOV, E. N.; FEDTCHOUK, A. S.; DYACHENKO, N. S.; NOSACH, L. N.; GRIDINA, T. L.; SHITIKOVA, L. I.; MUDRIK, L. M.; MESCHERIAKOV, A. K.; CHELOMBITKO, V. A.; ZHELTVAI, A. I.; EYNDE, J. V. The analysis of structure-anticancer and antiviral activity relationships for macrocyclic pyridinophanes and their analogues on the basis of 4D QSAR models (Simplex Representation of Molecular Structure). **Acta Biochimica Polonica**, Warsaw, v. 49, n. 1, p. 157–168, 2002.
117. MURATOV, E. N.; ARTEMENKO, A. G.; VARLAMOVA, E. V.; POLISCHUK, P. G.; LOZITSKY, V. P.; FEDCHUK, A. S.; REGINA L LOZITSKA, R. L.; TAT'YANA L GRIDINA, T. L.; KOROLEVA, L. S.; SIL'NIKOV, V. N.; GALABOV, A. S.; MAKAROV, V. A.; RIABOVA, O. B.; WUTZLER, P.; SCHMIDTKE, M.; KUZ'MIN, V. E. *Per aspera ad astra* : application of Simplex QSAR approach in antiviral research. **Future Medicinal Chemistry**, London, v. 2, n. 7, p. 1205-1226, 2010.
118. CARDOZO, M. G.; KAWAI, T.; IIMURA, Y.; SUGIMOTO, H.; YAMANISHI, Y.; HOPFINGER, A. J. Conformational analyses and molecular-shape comparisons of a series of indanone-benzylpiperidine inhibitors of acetylcholinesterase. **Journal of Medicinal Chemistry**, Washington, v. 35, n. 3, p. 590-601, 1992.
119. WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory System**, Amsterdam, v. 58, n. 2, p. 109-130, 2001.
120. CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, Washington, v. 110, n. 18, p. 5959-5967, 1988.
121. BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. **Journal of Microbiological Methods**, Amsterdam, v. 43, p. 3-31. 2000.

- 122 NIROGI, R.; SHINDE, A.; KAMBHAMPATI, R. S.; MOHAMMED, A. R.; SARAF, S. K.; BADANGE, R. K.; BANDYALA, T. R.; BHATTA, V.; BOJJA, K.; REBALLI, V.; SUBRAMANIAN, R.; BENADE, V.; PALACHARLA, R. C.; BHYRAPUNENI, G.; JAYARAJAN, P.; GOYAL, V.; JASTI, V. Discovery and Development of 1-[(2-Bromophenyl)sulfonyl]-5-methoxy-3-[(4-methyl-1-piperazinyl)methyl]-1 *H*-indole Dimesylate Monohydrate (SUVN-502): A Novel, Potent, Selective and Orally Active Serotonin 6 (5-HT<sub>6</sub>) Receptor Antagonist for Potential Treatment of Alzheimer's Disease. **Journal of Medicinal Chemistry**, Washington, v. 60, n. 5, p. 1843-1859, 2017.
- 123 SAYERS, E. W.; AGARWALA, R.; BOLTON, E. E.; BRISTER, J. R.; CANESE, K.; CLARK, K.; CONNOR, R.; FIORINI, N.; FUNK, K.; HEFFERON, T.; HOLMES, J. B.; KIM, S.; KIMCHI, A.; KITTS, P. A.; LATHROP, S.; LU, Z.; MADDEN, T. L.; MARCHLER-BAUER, A.; PHAN, L.; SCHNEIDER, V. A.; SCHOCH, C. L.; PRUITT, K. D.; OSTELL, J. Database resources of the National Center for Biotechnology Information. **Nucleic Acids Research**, Oxford, v. 47, p. 23-28, 2019.
- 124 HEO, L.; PARK, H.; SEOK, C. GalaxyRefine: protein structure refinement driven by side-chain repacking. **Nucleic Acids Research**, Oxford, v. 41, p. 384-388, 2013.
- 125 LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, Chichester, v. 26, p. 283-291, 1992.
- 126 SENA, M. M.; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, São Paulo, v. 23, n. 4, p. 547-559, 2000.
- 127 INFOMETRIX. **PIROUETTE**. Versão 3.10. [S. l.], 2002. Disponível em: <https://infometrix.com/pirouette/>. Acesso em: 27 ago. 2017.
- 128 FERREIRA, M. M. C.; MONTANARI, C. A.; GAUDIO, A. C. Seleção de variáveis em QSAR. **Química Nova**, São Paulo, v. 25, n. 3, p. 439-448, 2002.
- 129 LO, Y.; RENSI, S. E.; TORNG, W.; ALTMAN, R. B. Machine learning in chemoinformatics and drug discovery. **Drug Discovery Today**, London, v. 23, n. 8, p. 1538-1546, 2018.
- 130 MA, J.; SHERIDAN, R. P.; LIAW, A.; DAHL, G. E.; SVETNIK, V. Deep neural nets as a method for Quantitative Structure–Activity Relationships. **Journal of Chemical Information and Modeling**, Washington, v. 25, n. 2, p. 263-274, 2015.
- 131 AGATONOVIC-KUSTRIN, S.; BERESFORD, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. **Journal of Pharmaceutical and Biomedical Analysis**, Amsterdam, v. 22, n. 5, p. 717-727, 2000.
- 132 SHEELA, K. G.; DEEPA, S. N. Review on methods to fix number of hidden neurons in neural networks. **Mathematical Problems in Engineering**, London, v. 2013, p. 1-12, 2013.

133 LILLICRAP, T. P.; SANTORO, A.; MARRIS, L.; AKERMAN, C. J.; HINTON, G. Backpropagation and the brain. **Nature Reviews Neuroscience**, London, v. 21, n. 6, p. 335-346, 2020.

134 JALALI-HERAVI, M.; ASADOLLAHI-BABOLI, M.; SHAHBAZIKHAH, P. QSAR study of heparanase inhibitors activity using artificial neural networks and Levenberg–Marquardt algorithm. **European Journal of Medicinal Chemistry**, Issy les Moulineaux Cedex, v. 43, n. 3, p. 548-556, 2007.

135 KHAN, N.; GAURAV, D.; KANDL, T. Performance evaluation of levenberg-marquardt technique in error reduction for diabetes condition classification. **Procedia Computer Science**, Amsterdam, v. 18, p. 2629-2637, 2013.

136 DA SILVA, A. P.; DE ANGELO, R. M.; DE PAULA, H.; HONÓRIO, K. M.; DA SILVA, A. B. F. Drug design of new 5-HT<sub>6</sub> antagonists: a QSAR study of arylsulfonamide derivatives. **Structural Chemistry**, New York, v. 32, p. 1585-1597, 2020.

137 DA SILVA, A. P.; CHIARI, L. P. A.; GUIMARAES, A. R.; HONORIO, K. M.; DA SILVA, A. B. F. Drug design of new 5-HT<sub>6R</sub> antagonists aided by artificial neural networks. **Journal of Molecular Graphic and Modelling**, Philadelphia, v. 104, n. 107844, 2021.

138 SCHRÖDINGER. **Maestro- Schrödinger**. Versão 2018-1. New York, 2018. Disponível em: <https://www.schrodinger.com/releases/release-2018-1>. Acessado em: 5 jan. 2018.

139 PHILLIPS, J. C.; HARDY, D. J.; MAIA, J. D. C.; STONE, J. E.; RIBEIRO, J. V.; BERNARDI, R. C.; BUCH, R.; FIORIN, G.; HÉNIN, J.; JIANG, W.; MCGREEVY, R.; MELO, M. C. R.; RADAK, B. K.; SKEEL, R. D.; SINGHARROY, A.; WANG, Y.; ROUX, B.; AKSIMENTIEV, A.; LUTHEY-SCHULTEN, Z.; KALÉ, L. V.; SCHULTEN, K.; CHIPOT, C.; TAJKHORSHID, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. **The Journal of Chemical Physics**, Melville, v. 153, n. 4, p. 044130-044133, 2020.

140 FELLER, S. E.; ZHANG, Y.; PASTOR, R. W.; BROOKS, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. **The Journal of Chemical Physics**, Melville, v. 103, n. 4613, p. 4313-4621, 1995.

141 ESSMANN, U.; PERERA, L.; BERKOWITZ, M. L.; DARDEN, T.; LEE, H.; PEDERSEN, L. G. A smooth particle mesh Ewald method. **The Journal of Chemical Physics**, Melville, v. 103, n. 8577, p. 8577-8593, 1995.

142 YU, W.; HE, X.; VANOMMESLAEGHE, K.; MACKERELL, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. **Journal of Computational Chemistry**, Hoboken, v. 33, n. 31, p. 2451-2468, 2012.

143 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphic and Modelling**, Philadelphia, v.14, n. 1, p. 33-38, 1996.

144 FRISCH, M. J.; TRUCKS, G. W.; SCHLEGEL, H. B.; SCUSERIA, G. E.; ROBB, M. A.; CHEESEMAN, J. R.; SCALMANI, G.; BARONE, V.; PETERSSON, G. A.; NAKATSUJI, H.; LI, X.; CARICATO, M.; MARENICH, A.; BLOINO, J.; JANESKO, B. G.; GOMPERTS, R.; MENNUCCI, B.; HRATCHIAN, H. P.; ORTIZ, J. V.; IZMAYLOV, A. F.; SONNENBERG, J. L.; WILLIAMS-YOUNG, D.; DING, F.; LIPPARINI, F.; EGIDI, F.; GOINGS, J.; PENG, B.; PETRONE, A.; HENDERSON, T.; RANASINGHE, D.; ZAKRZEWSKI, V. G.; GAO, J.; REGA, N.; ZHENG, G.; LIANG, W.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; VREVEN, T.; THROSSELL, K.; MONTGOMERY JUNIOR, J. A.; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M.; HEYD, J. J.; BROTHERS, E.; KUDIN, K. N.; STAROVEROV, V. N.; KEITH, T.; KOBAYASHI, R.; NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A.; BURANT, J. C.S.; IYENGAR, S.; TOMASI, J.; COSSI, M.; MILLAM, J. M.; KLENE, M.; ADAMO, C.; CAMMI, R.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; FARKAS, O.; FORESMAN, J. B.; FOX, D. J. **Gaussian 09**: revision A.02. Wallingford: Gaussian, 2016.

145 LEE, C.; YANG, W.; PARR, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Physical Review B**, College Park, v. 37, n. 2, p. 785-789, 1988.

146 DITCHFIELD, R.; HEHRE, W. J.; POPLE, J. A. Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules. **The Journal of Chemical Physics**, Melville, v. 54, n. 2, p. 724-728, 1971.

147 TETKO, I. V.; GASTEIGER, J.; TODESCHINI, R.; MAURI, A.; LIVINGSTONE, D.; ERTL, P.; PALYULIN, V. A.; RADCHENKO, E. V.; ZEFIROV, N. S.; MAKARENKO, A. S. TANCHUK, V. Y.; PROKOPENKO, V. V. Virtual computational chemistry laboratory – design and description. **Journal of Computer-Aided Molecular Design**, Dordrecht, v. 19, n. 6, p. 453–463, 2005.

148 AGUIAR-PULIDO, V.; GESTAL, M.; CRUZ-MONTEAGUDO, M.; RABUNAL, J.; DORADO, J.; MUNTEANU, C. Evolutionary computation and QSAR research. **Current Computer-Aided Drug Design**, Sharjah, v. 9, n. 2, p. 206-225. 2013.

149 **SYBYL**. Versão 8.1. St. Louis: Tripos International, 2008.

150 MATHWORKS. **MATLAB**. Versão 2018. [S. l.], 2018. Disponível em: <https://www.mathworks.com/products/matlab.html>. Acesso em: 12 ago. 2018

151 **SCHRÖDINGER Release 2021-1. QikProp**. New York, 2021. Disponível em: <https://www.schrodinger.com/products/qikprop>. Acesso em: 20 jan. 2021.

152 DAINA, A.; MICHIELIN, O.; ZOETE, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. **Scientific Report**, London, v. 7, n. 42717, p. 1-13, 2017.

153 NUGENT, T.; JONES, D. T. Transmembrane protein topology prediction using support vector machines. **BMC Bioinformatics**, London, v. 10, n. 159, p. 1-11, 2009.

- 154 SHEN, M.; SALI, A. Statistical potential for assessment and prediction of protein structures. **Protein Science**, Hoboken, v. 15, n. 11, p. 2507-2524, 2016.
- 155 TETKO, I. V.; GASTEIGER, J.; TODESCHINI, R.; MAURI, A.; LIVINGSTONE, D.; ERTL, P.; PALYULIN, V. A.; RADCHENKO, E. V.; ZEFIROV, N. S.; MAKARENKO, A. S.; TANCHUK, V. Y.; PROKOPENKO, V. V. Virtual computational chemistry laboratory--design and description. Virtual computational chemistry laboratory--design and description. **Journal of Computer-Aided Molecular Design**, Dordrecht, v. 19, n. 6, 453-463, 2005.
- 156 TROPSHA, A. GRAMATICA, P.; GOMBAR, V. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. **QSAR & Combinatorial Science**, Weinheim, v. 22, n. 1, p. 69-77, 2003.
- 157 GOLBRAIKH, A.; TROPSHA, A. Beware of q2! **Journal of Molecular Graphics and Modelling**, Philadelphia, v. 20, n. 4, p. 269-276, 2002.
- 158 KIRALJ, R.; FERREIRA, M. M. C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. **Journal of Brazilian Chemical Society**, São Paulo, v. 20, n. 4, p. 770-787, 2009.
- 159 ZARE-SHAHABADI, V.; LOTFIZADEH, M.; GANDOMANI, A. R. A.; PAPARI, M. M. Determination of boiling points of azeotropic mixtures using quantitative structure-property relationship (QSPR) strategy. **Journal Molecular Liquids**, Amsterdam, v. 188, p. 222-229, 2013.
- 160 BURDEN, F. R. Molecular identification number for substructure searches. **The Journal of Chemical Information and Computer Sciences**, Washington, v. 29, p. 255-227, 1989.
- 161 HEMMER, M. C.; STEINHAEUER, V.; GASTEIGER, J. Deriving the 3D structure of organic molecules from their infrared spectra. **Vibrational Spectroscopy**, Amsterdam, v. 19, p. 151-164, 1998.
- 162 CLARE, B. W. Charge transfer complexes and frontier orbital energies in QSAR: a congeneric series of electron acceptors. **Journal of Molecular Structure: THEOCHEM**, Amsterdam, v. 337, n. 2, p. 139-150, 1995.
- 163 HEATON C. A.; MILLER, A. K.; POWELL, R. L. Predicting the reactivity of fluorinated compounds with copper using semi-empirical calculations. **Journal of Fluorine Chemistry**, Amsterdam, v. 107, n. 1, p. 1-3, 2011.
- 164 HONÓRIO K. M.; DA SILVA, A. B. F. An AM1 study on the electron-donating and electron-accepting character of biomolecules. **International Journal of Quantum Chemistry**, Hoboken, v. 95, n. 2, p. 126-132, 2003.
- 165 TODESCHINI, R.; CONSONNI, V. **Handbook of molecular descriptors**. Weinheim: Wiley-VCH, 2000. 668 p. ISBN:9783527299133.

166 CONSONNI, V.; TODESCHINI, R.; PAVAN, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. **Journal of Chemical Information and Computer Sciences**, Washington, v. 42, n. 3, p. 682-692, 2002.

167 SCHWARTZ, T. W.; FRIMURER, T. M.; HOLST, B.; ROSENKILDE, M. M.; ELLING, C. E. Molecular mechanism of 7TM receptor activation—a global toggle switch model. **Annual Review Pharmacology and Toxicology**, Palo Alto, v. 46, n. 1, p. 481-519, 2016.

168 ZHUANG W.; WANG Y, CUI P-F, XING L, LEE J, KIM D, JIANG HL, OH YK. Applications of  $\pi$ - $\pi$  stacking interactions in the design of drug-delivery systems. **Journal of Controlled Release**, Amsterdam, v. 294, p. 311-326, 2019.

169 CAVALLO G.; METRANGOLO P.; MILANI R.; PILATI T.; PRIIMAGI A.; RESNATI G.; TERRANEO, G. The halogen bond. **Chemical Reviews**, Washington, v. 166, n. 4, p. 2478-2601, 2016.

170 FORD M. C.; HO, P. S. Computational tools to model halogen bonds in medicinal chemistry. **Journal of Medicinal Chemistry**, Washington, v. 59, n. 5, p. 1655-1670, 2016.

171 LARREGIEU C.A.; BENET, L. Z. Drug discovery and regulatory considerations for improving in silico and in vitro predictions that use caco-2 as a surrogate for human intestinal permeability measurements. **The AAPS Journal**, New York, v. 15, n. 2, p. 483-497, 2013.

## APÊNDICE A

**Tabela A1** - Valores dos descritores para cada composto do conjunto **1** (compostos do conjunto teste está em **negrito e sublinhado**)

Composto	X5sol	BELm3	MATS6p	MATS1v	RDF140m	GATS8e	E <sub>lumo</sub> (eV)*
C1	7,53	1,08	-0,19	0,031	0,00	2,235	-1,082
C2	8,21	1,08	-0,20	0,03	0,00	2,175	-1,46
C3	7,97	1,08	-0,27	0,029	0,00	2,489	-0,844
C4	8,06	1,08	-0,12	-0,005	0,003	2,913	-0,840
C5	9,85	1,08	-0,13	0,045	0,00	2,449	-1,542
C6	7,61	1,82	-0,004	-0,012	0,085	3,066	-1,220
C7	7,33	1,82	-0,007	-0,013	0,00	2,491	-1,288
C8	7,35	1,82	-0,005	-0,013	0,002	2,863	-1,330
C9	7,03	1,82	-0,003	-0,013	0,006	2,441	-1,273
C10	8,00	1,82	0,037	-0,013	0,00	2,801	-1,379
C11	8,39	1,82	0,011	-0,013	0,003	3,034	-1,382
C12	8,04	1,82	-0,026	-0,013	0,041	2,761	-1,372
C13	8,23	1,82	-0,093	-0,043	0,00	3,042	-1,151
C14	7,09	1,82	0,054	-0,043	0,025	3,401	-1,099
C15	8,05	1,83	0,05	-0,011	0,019	3,113	-1,218
C16	7,80	1,83	-0,099	-0,011	0,007	3,167	-1,165
C17	7,49	1,58	-0,009	-0,007	0,00	3,143	-1,313
C18	8,28	1,82	0,072	-0,013	0,003	2,622	-1,358
C19	9,68	1,825	0,023	0,008	0,011	3,017	-1,550
C20	9,45	1,802	0,016	0,045	0,037	3,237	-1,595
C21	7,13	1,08	-0,018	0,031	0,115	2,339	-1,244
C22	7,82	1,08	-0,025	0,03	1,105	1,714	-1,304
C23	7,58	1,801	-0,12	0,029	0,685	1,935	-1,043
C24	7,67	1,08	0,039	-0,005	1,698	2,753	-1,047
C25	7,33	1,08	-0,135	0,029	3,323	1,337	-1,509
C26	9,46	1,801	0,023	0,045	0,146	1,008	-1,495
C27	9,84	1,824	0,013	0,008	0,00	3,079	-1,510
C28	8,01	1,806	0,055	-0,012	0,00	2,907	-1,121
C29	8,24	1,814	-0,069	-0,011	0,00	2,959	-1,168
C30	10,31	1,083	0,079	-0,024	0,011	2,741	-1,518
C31	10,08	1,807	0,075	0,008	0,007	2,894	-1,486
C32	10,54	1,835	-0,027	-0,022	0,249	2,813	-1,459
C33	10,31	1,814	-0,036	0,009	0,015	2,954	-1,478
C34	10,72	1,844	-0,028	-0,02	0,008	3,403	-1,323
C35	10,50	1,827	-0,037	0,009	0,004	3,543	-1,416
C36	10,74	1,847	-0,114	-0,02	0,024	2,886	-1,442
C37	10,51	1,831	-0,126	0,009	0,002	3,018	-1,468
C38	10,86	1,867	-0,026	-0,019	0,234	3,923	-1,508
C39	10,64	1,086	-0,037	0,009	0,00	4,057	-1,343
C40	11,99	1,956	0,1	-0,007	0,11	3,361	-1,373
C41	11,77	1,954	0,096	0,021	0,011	3,005	-1,486
C42	6,93	1,803	-0,061	0,025	0,00	2,438	-0,957
C43	7,21	1,803	-0,022	0,023	0,00	2,335	-1,180
C44	7,40	1,803	0,000	0,021	0,00	2,407	-1,146
C45	7,75	1,803	-0,071	0,023	0,00	2,448	-1,229
C46	7,04	1,803	0,022	0,024	0,00	1,586	-1,425
C47	7,59	1,802	-0,127	0,021	0,00	2,344	-1,266
C48	7,02	1,803	-0,004	-0,005	0,00	3,028	-0,724
C49	6,86	1,803	-0,172	0,024	0,00	1,665	-1,487
C50	7,08	1,803	0,002	-0,006	0,00	1,104	-1,116
C51	7,29	1,803	0,000	0,022	0,00	4,695	-0,879
C52	8,68	1,803	-0,105	0,039	0,478	2,837	-1,475
C53	10,61	1,803	-0,06	0,04	0,008	2,157	-1,582
C54	6,88	1,802	-0,018	0,023	0,00	1,935	-1,103
C55	8,12	1,802	-0,122	0,022	0,492	2,278	-1,429
C56	8,69	1,802	0,022	-0,006	0,00	2,253	-1,022
C57	7,05	1,797	-0,035	0,063	0,00	2,663	-1,205
C58	7,03	1,797	-0,098	0,058	0,00	2,551	-1,300
C59	6,11	1,797	-0,027	0,066	0,00	2,039	-1,039

\*Eletrovolt

## APÊNDICE B

**Tabela B1-Valores dos descritores para cada um dos compostos do conjunto 2.**

Composto	pK <sub>i</sub>	JhetZ	BEHm2	GATS3p	GATS1p	RDF025u	Au	H3v	E <sub>HOMO-2</sub>
1	8,22	2,00	3,00	2,00	2,00	38,00	83,00	1,00	-0,22
2	7,99	2,00	4,00	2,00	2,00	37,00	83,00	1,00	-0,22
3	7,72	2,00	4,00	2,00	2,00	36,00	84,00	1,00	-0,22
4	8,67	2,00	4,00	2,00	2,00	37,00	82,00	1,00	-0,22
5	7,79	2,00	3,00	2,00	2,00	37,00	83,00	1,00	-0,22
6	8,11	2,00	3,00	2,00	2,00	39,00	82,00	1,00	-0,22
7	7,24	2,00	3,00	2,00	2,00	37,00	84,00	1,00	-0,22
8	8,06	2,00	3,00	2,00	2,00	38,00	82,00	1,00	-0,22
9	8,22	2,00	3,00	2,00	2,00	48,00	124,00	1,00	-0,22
10	8,02	2,00	3,00	2,00	2,00	40,00	99,00	1,00	-0,22
11	7,53	2,00	3,00	2,00	2,00	39,00	108,00	1,00	-0,22
12	7,60	2,00	3,00	2,00	2,00	40,00	101,00	1,00	-0,22
13	6,82	2,00	3,00	2,00	2,00	39,00	104,00	1,00	-0,22
14	7,04	2,00	3,00	2,00	2,00	39,00	88,00	1,00	-0,22
15	7,52	2,00	3,00	2,00	2,00	37,00	83,00	1,00	-0,22
16	6,92	2,00	3,00	2,00	2,00	36,00	83,00	1,00	-0,22
17	7,19	2,00	3,00	2,00	2,00	38,00	82,00	1,00	-0,22
18	7,55	2,00	3,00	2,00	2,00	35,00	88,00	1,00	-0,22
19	7,00	2,00	3,00	1,00	2,00	35,00	87,00	1,00	-0,22
20	7,54	2,00	3,00	2,00	2,00	36,00	87,00	1,00	-0,22
21	7,04	2,00	3,00	2,00	2,00	41,00	111,00	1,00	-0,22
22	7,19	2,00	3,00	2,00	2,00	40,00	99,00	1,00	-0,22
23	7,94	2,00	3,00	2,00	2,00	46,00	105,00	1,00	-0,22
24	8,04	2,00	4,00	2,00	2,00	46,00	130,00	1,00	-0,22
25	7,51	2,00	4,00	2,00	2,00	37,00	83,00	1,00	-0,22
26	7,63	2,00	4,00	1,00	2,00	36,00	86,00	1,00	-0,22
27	7,76	2,00	4,00	2,00	2,00	36,00	85,00	1,00	-0,23
28	7,45	2,00	4,00	2,00	2,00	37,00	85,00	1,00	-0,22
29	7,22	2,00	4,00	2,00	2,00	35,00	84,00	1,00	-0,22
30	8,40	2,00	3,00	2,00	2,00	47,00	120,00	1,00	-0,22
31	7,92	2,00	4,00	2,00	2,00	36,00	83,00	1,00	-0,22
32	8,04	2,00	4,00	2,00	2,00	36,00	84,00	1,00	-0,23
33	7,77	2,00	4,00	2,00	2,00	36,00	82,00	1,00	-0,23
34	7,56	2,00	4,00	2,00	2,00	46,00	122,00	1,00	-0,22
35	7,94	2,00	3,00	2,00	2,00	50,00	152,00	1,00	-0,21
36	7,84	2,00	4,00	2,00	2,00	38,00	95,00	1,00	-0,21
37	7,74	2,00	3,00	2,00	2,00	39,00	102,00	1,00	-0,22
38	8,70	2,00	4,00	2,00	2,00	38,00	97,00	1,00	-0,21
39	8,79	2,00	3,00	2,00	2,00	49,00	138,00	1,00	-0,22
40	8,30	2,00	3,00	2,00	2,00	39,00	96,00	1,00	-0,22
41	8,37	2,00	3,00	2,00	2,00	39,00	96,00	1,00	-0,21

**Tabela B1 - Continuação**

42	7,99	2,00	4,00	2,00	2,00	38,00	98,00	1,00	-0,22
43	8,11	2,00	3,00	2,00	2,00	47,00	118,00	1,00	-0,21
44	7,81	2,00	3,00	2,00	2,00	42,00	115,00	1,00	-0,21
45	7,74	2,00	3,00	2,00	2,00	41,00	121,00	1,00	-0,21
46	7,42	2,00	3,00	2,00	2,00	41,00	122,00	1,00	-0,22
47	7,44	2,00	4,00	1,00	2,00	45,00	120,00	1,00	-0,21
48	7,02	2,00	4,00	2,00	2,00	40,00	122,00	1,00	-0,22
49	7,44	2,00	3,00	2,00	2,00	49,00	147,00	1,00	-0,22
50	6,62	2,00	3,00	2,00	2,00	39,00	108,00	1,00	-0,22
51	7,73	2,00	4,00	2,00	2,00	37,00	88,00	1,00	-0,21
52	7,30	2,00	3,00	2,00	2,00	48,00	130,00	1,00	-0,22
53	7,46	2,00	3,00	2,00	2,00	38,00	87,00	1,00	-0,22
54	6,70	2,00	4,00	2,00	2,00	37,00	94,00	1,00	-0,23
55	6,71	2,00	4,00	1,00	2,00	37,00	99,00	1,00	-0,22
56	7,25	2,00	3,00	2,00	2,00	38,00	86,00	1,00	-0,22
57	7,53	2,00	4,00	2,00	2,00	47,00	133,00	1,00	-0,22
58	8,24	2,00	4,00	1,00	2,00	41,00	93,00	1,00	-0,21
59	8,70	2,00	3,00	2,00	2,00	52,00	122,00	1,00	-0,21
60	8,11	2,00	3,00	2,00	2,00	42,00	79,00	1,00	-0,21
61	8,17	2,00	4,00	1,00	2,00	43,00	111,00	1,00	-0,21
62	8,77	2,00	3,00	2,00	2,00	54,00	154,00	1,00	-0,21
63	7,70	2,00	3,00	2,00	2,00	44,00	109,00	1,00	-0,22
64	8,06	2,00	3,00	2,00	2,00	44,00	109,00	1,00	-0,22
65	6,51	1,00	4,00	2,00	1,00	39,00	81,00	1,00	-0,20
66	6,49	2,00	4,00	1,00	1,00	39,00	71,00	1,00	-0,22
67	6,32	1,00	3,00	2,00	2,00	28,00	33,00	0,00	-0,22
68	8,12	2,00	3,00	2,00	2,00	37,00	71,00	1,00	-0,22
69	8,35	2,00	4,00	1,00	2,00	36,00	71,00	1,00	-0,22
70	8,89	2,00	4,00	1,00	2,00	37,00	86,00	1,00	-0,22
71	8,24	2,00	3,00	2,00	2,00	48,00	122,00	1,00	-0,22
72	8,16	2,00	3,00	2,00	2,00	38,00	84,00	1,00	-0,22
73	6,44	1,00	4,00	1,00	2,00	48,00	99,00	1,00	-0,24
74	6,41	2,00	4,00	1,00	2,00	39,00	71,00	1,00	-0,24
75	6,88	1,00	4,00	2,00	2,00	47,00	137,00	1,00	-0,22
76	6,46	2,00	4,00	1,00	2,00	42,00	111,00	1,00	-0,22
77	6,40	2,00	6,00	1,00	1,00	37,00	74,00	1,00	-0,23
78	6,54	2,00	6,00	1,00	1,00	36,00	82,00	1,00	-0,23
79	7,02	2,00	4,00	1,00	2,00	35,00	73,00	1,00	-0,22
80	6,56	2,00	3,00	2,00	2,00	36,00	75,00	0,00	-0,22
81	6,46	2,00	4,00	1,00	2,00	37,00	108,00	1,00	-0,22