Laise Pellegrini Alencar Chiari

Estudos de Relações Quantitativas Estrutura-Atividade de Antagonistas do Receptor Sigma-1

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

Área de concentração: Físico-química

Orientador: Prof. Dr. Albérico Borges
Ferreira da Silva

São Carlos, SP

2017

*Aos meus pais Valdir e Rita, meu marido Thiago e
minhas preciosas filhas, Manuela e Melissa.*

AGRADECIMENTOS

Primeiramente agradeço pelo imenso, eterno e perfeito amor de meu Pai Celestial. Por Seu Plano de Salvação, por ter me concedido a vida e ter me provido um Salvador, me dando inúmeras oportunidades de aprendizado e crescimento nesta Terra. Por ouvir minhas orações e, especialmente, por respondê-las sempre. Por me dar a oportunidade de conhecer seu Evangelho restaurado e assim receber diariamente as bênçãos que provêm dele.

Ao meu Salvador Jesus Cristo. Pelo seu exemplo único, por sua humildade, pelo seu amor sublime. Pela sua Expição, seu poder redentor e por poder me amar e compreender como ninguém jamais o faria.

Ao meu querido pai Valdir. Por ser meu maior exemplo de trabalho e retidão. Por ter trabalhado tanto para nos dar o melhor que podia, mas por nunca ter deixado de dedicar seu tempo para nós e se fazer sempre presente, deixando inúmeras recordações em família que guardo com tanto carinho.

À minha querida mãe Rita. Por ser tão solícita, amável e compreensiva. Por ter sempre dado atenção a cada detalhe de minhas necessidades e por fazê-las parecer sempre importantes, mesmo que fossem pequenas. Por seu companheirismo e carinho. Por seu grande coração e por emanar tanto amor. Por ser além de minha mãe, minha melhor amiga.

Aos meus familiares, avós, tios, primos e sogros, pelo carinho sempre.

A todos os amigos que fizeram e fazem parte da minha história. Em especial a Débora e Juliana, que são amigas especiais que se fizeram presentes em muitos momentos da minha vida tornando-os mais felizes. À querida Tatiana, por ter me ajudado nos momentos mais difíceis e pelos seus sábios conselhos.

Ao professor Albérico, que me incentivou a continuar os estudos, por seu desejo de sempre ajudar, por seu bom humor contagiante e pelas conversas que sempre rendem boas risadas.

A todos os colegas do laboratório, que acolheram não só a mim, mas também a minha pequena Melissa. Agradeço em especial aos amigos Célio e Aline, sem os quais eu jamais teria conseguido realizar esse trabalho. Por serem sempre tão solícitos e pacientes ao me ensinar e ajudar. E à querida Aline, pelo seu exemplo de luta e determinação. Pelo seu sorriso em meio à sua provação. Obrigada por seu exemplo de força e perseverança.

Ao meu marido e amor da minha vida, Thiago. Pelo companheirismo, compreensão, carinho. Pelo apoio e incentivo, sempre. Por compartilhar sua vida comigo e me fazer uma pessoa melhor, mais feliz e completa. Pelos momentos que já passamos de alegrias e de dificuldades e pelos muitos

momentos que ainda estão por vir. Que a eternidade juntos seja a maior bênção que possamos receber.

Às minhas preciosas filhas, Manuela e Melissa. Por serem minhas companheirinhas, adicionando graça, beleza e alegria todos os dias da minha vida como ninguém jamais o faria. Por me concederem os momentos e as recordações mais doces e belas. Por me darem forças para ser uma pessoa melhor em todos os aspectos. Por serem o que tenho de mais precioso. Meu amor por vocês é maior do que tudo o que podem imaginar. E eu espero que nossa família possa ser eterna.

*“O futuro é tão brilhante quanto à sua fé.”
(Thomas S. Monson)*

RESUMO

A dor neuropática atinge cerca de 6 a 10% da população global e estima-se o seu aumento nos próximos anos. Essa síndrome não tem cura e afeta consideravelmente a qualidade de vida das pessoas por ela acometidas. Os medicamentos utilizados atualmente para o seu tratamento, como antidepressivos, anticonvulsivantes, opióides, dentre outros, não proporcionam um resultado satisfatório pelo fato de não reduzirem consideravelmente os sintomas e/ou por terem muitos efeitos colaterais. Pesquisas recentes mostram que o receptor sigma-1 pode ser utilizado no tratamento da dor neuropática. Verificou-se na literatura uma nova série de pirimidinas que são capazes de se ligar ao receptor sigma-1, atuando como seus antagonistas, sendo potenciais alvos para a produção de fármacos que podem ser utilizados no tratamento da dor neuropática. Então, estudos de Relações Quantitativas Estrutura-Atividade (QSAR) foram realizados utilizando os métodos de Mínimos Quadrados Parciais (PLS) e Redes Neurais Artificiais (ANN) para prever a atividade biológica dessa série de pirimidinas. Os resultados obtidos se mostraram satisfatórios tanto para o método de PLS ($r^2 = 0,877$, $q^2 = 0,800$ e $r^2_{\text{teste}} = 0,738$), quanto para o método de ANN ($r^2_{\text{trein}} = 0,734$, $r^2_{\text{val}} = 0,753$ e $r^2_{\text{teste}} = 0,676$), mostrando que o conjunto de compostos antagonistas do receptor sigma-1 pode ser descrito tanto de forma linear quanto de forma não-linear.

Palavras-chave: Dor neuropática, receptor Sigma-1, pirimidinas, QSAR, PLS, MLP-ANN.

ABSTRACT

Neuropathic pain affects about 6 to 10% of the global population and it is estimated to increase in the coming years. This syndrome has no cure and considerably affects the life quality of people affected by it. Medications currently used for its treatment, such as antidepressants, anticonvulsants, opioids, among others, do not provide a satisfactory result because they do not significantly reduce the symptoms and/or have many side effects. Recent research shows that the sigma-1 receptor can be used in the treatment of the neuropathic pain. A new series of pyrimidines have been found in the literature, which are capable of binding to the sigma-1 receptor, acting as its antagonists, and have been synthesized as potential targets that can be used in the treatment of the neuropathic pain. Therefore, Quantitative Structure-Activity Relationships (QSAR) were performed using Partial Least Squares (PLS) and Artificial Neural Networks (ANN) methods to predict the biological activity of this series of pyrimidines. Through the mathematical models obtained by PLS ($r^2 = 0.877$, $q^2 = 0.800$ and $r^2_{\text{test}} = 0.738$) and ANN ($r^2_{\text{train}} = 0.734$, $r^2_{\text{val}} = 0.753$ and $r^2_{\text{test}} = 0.676$) methods, it was showed that they were able to predict the biological activity of the studied pyrimidines.

Keywords: Neuropathic pain, Sigma-1R, pyrimidines, QSAR, PLS, MLP-ANN.

LISTA DE FIGURAS

Figura 1. Estrutura geral de um composto antagonista do receptor Sigma-1...	16
Figura 2. Estruturas químicas de compostos antagonistas do receptor Sigma-1 que podem atuar no tratamento da dor neuropática.....	16
Figura 3. Estrutura molecular do composto de referência formada pelo esqueleto de pirimidina.....	17
Figura 4. Esquema geral do funcionamento da técnica de algoritmo genético.....	19
Figura 5. Esquema geral do funcionamento de uma MLP-ANN.....	23
Figura 6. Distribuição dos valores de pK_i para a série de pirimidinas, antagonistas do receptor sigma-1.....	25
Figura 7. Estruturas moleculares dos compostos 1 ao 6.....	26
Figura 8. Estruturas moleculares dos compostos 7 ao 10.....	27
Figura 9. Estruturas moleculares dos compostos 11 ao 16.....	28
Figura 10. Estrutura molecular do composto 17.....	29
Figura 11. Estruturas moleculares dos compostos 19 a 20.....	29
Figura 12. Estruturas moleculares dos compostos 21 a 27 e 31 a 37.....	30
Figura 13. Estruturas moleculares dos compostos 28 a 30 e 38 a 40.....	31
Figura 14. Estruturas moleculares dos compostos 41 a 43.....	32
Figura 15. Estruturas moleculares dos compostos 44 ao 50.....	32
Figura 16. Estruturas moleculares dos compostos 51 a 54.....	33
Figura 17. Coeficientes de regressão para o modelo PLS.....	43
Figura 18. Detecção de outliers no modelo PLS.....	44

Figura 19. Estruturas moleculares dos compostos que apresentaram altos valores de alavancagem.....	45
Figura 20. Estrutura molecular do composto que apresentou valor fora do limite de resíduos de Student.....	45
Figura 21. Testes de <i>leave-N-out</i> aplicados ao modelo PLS.....	46
Figura 22. Testes de randomização aplicados ao modelo PLS.....	47
Figura 23. Predição da atividade biológica pelo modelo PLS.....	48
Figura 24. Estrutura molecular do composto 52.....	49
Figura 25. Desempenho das MLP-ANNs treinadas.....	50
Figura 26. Predição da atividade biológica pelo modelo ANN.....	52
Figura 27. Compostos que apresentaram valor igual a 2 para o descritor BELm1.....	54
Figura 28. Compostos com carbonos sp^3 na região hidrofóbica secundária..	55
Figura 29. Compostos com carbono sp^3 na posição R_3	55
Figura 30. Composto com carbono sp^3 na região hidrofóbica primária.....	56
Figura 31. Compostos que apresentaram baixos valores de E_{LUMO+4}	57
Figura 32. Compostos com baixo valor de pK_i e alto valor de E_{LUMO+4}	58
Figura 33. Compostos que apresentaram valor nulo para o descritor Mor12e.....	58
Figura 34. Compostos que apresentaram valor igual a 1 para o descritor J3D.....	59
Figura 35. Principais compostos do descritor momento dipolo.....	60
Figura 36. Compostos com alteração no tamanho da cadeia de carbonos que liga a pirimidina e a morfolina.....	61
Figura 37. Compostos com alterações na cadeia de carbonos.....	62

Figura 38. Compostos com alteração na morfolina.....	62
Figura 39. Compostos que apresentaram valor igual a 15 para o descritor T(N..O).....	63
Figura 40. Compostos com alterações nas posições R ₃ e na região hidrofóbica primária.....	63
Figura 41. Compostos que apresentaram valor igual a 3 para o descritor ClC ₀	64

LISTA DE TABELAS

Tabela 1. Parâmetros estatísticos do modelo gerado pela técnica de algoritmo genético.....	39
Tabela 2. Valores dos descritores selecionados para realizar a geração de modelos PLS e ANN.....	39
Tabela 3. Variância percentual acumulada, SEV, SEC, PRESS, q^2 e r^2 dos modelos PLS obtidos para o conjunto treinamento.....	42
Tabela 4. Resíduos de predição do conjunto teste pelo modelo PLS.....	48
Tabela 5. Melhores modelos de MLP-ANN gerados para cada combinação de função de transferência.....	51
Tabela 6. Resíduos de predição do conjunto teste pelo modelo ANN.....	52
Tabela 7. Descritores selecionados para os modelos QSAR.....	53

LISTA DE ABREVIATURAS E SIGLAS

3DMoRSE	<i>3D Molecule Representation of Structures based on Electron diffraction</i>
ANN	<i>Artificial Neural Network</i>
DFT	<i>Density Functional Theory</i>
DMA	Desvio Médio Absoluto
EQM	Erro Quadrático Médio
IASP	<i>International Association for the Study of Pain</i>
MAM	<i>Mitochondrion-associated ER membrane</i>
MLP-ANN	<i>Multilayer Perceptron Artificial Neural Network</i>
Mor12e	Sinal 12 dos descritores 3D-MoRSE ponderado pela eletronegatividade
PLS	<i>Partial Least Squares</i>
PRESS	<i>Prediction residues errors squares sum</i>
QSAR	<i>Quantitative Structure-Activity Relationships</i>
SEC	<i>Standard error of calibration</i>
SEV	<i>Standard error of validation</i>

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 Métodos Computacionais.....	17
1.1.1 Geração do Modelo QSAR - Método PLS.....	20
1.1.2 Geração do Modelo QSAR – Método ANN.....	22
2 OBJETIVOS.....	24
3 METODOLOGIA.....	25
3.1 Seleção de variáveis.....	35
3.2 Geração dos modelos QSAR.....	35
4 RESULTADOS E DISCUSSÃO.....	38
4.1 Seleção de variáveis.....	38
4.2 Geração de modelos QSAR.....	41
4.2.1 Mínimos Quadrados Parciais (PLS).....	41
4.2.2 Redes Neurais Artificiais (ANN).....	49
4.2.3 Descritores Seleccionados.....	53
5. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS.....	65
REFERÊNCIAS.....	66
APÊNDICE.....	76

1 INTRODUÇÃO

A dor pode ser de origem nociceptiva ou neuropática. A dor nociceptiva surge como consequência de estímulos prejudiciais aos tecidos ósseos, musculares ou ligamentares, tendo como principal função a de sinalizar esses danos. A dor neuropática surge por lesão ou disfunção do sistema nervoso tendo como causas mais frequentes a neuropatia diabética e neuropatias associadas com dor lombar crônica¹. No entanto, a dor neuropática pode não apresentar a princípio uma causa identificável e aparente^{2,3}.

De acordo com a Associação Internacional para o Estudo da Dor (do inglês IASP - *International Association for the Study of Pain*), a dor neuropática é definida como “dor causada por uma lesão ou doença do sistema somatosensorial”⁴. Ela afeta de 6 a 10% da população global, sendo reconhecida como uma das síndromes de dor mais complicadas de ser tratada pelos fatos de ser de difícil avaliação clínica e diagnóstica e devido aos medicamentos utilizados atualmente para o seu tratamento não trazerem resultados eficazes⁵. Estima-se o aumento da prevalência da dor neuropática nos próximos anos como consequência do aumento da sobrevida de pacientes com doenças crônicas associadas a este tipo de dor e ao envelhecimento populacional⁶.

Os sintomas apresentados pelas pessoas que sofrem com essa síndrome são inúmeros e variados, sendo que são classificados como positivos e negativos. Os sintomas negativos são desconfortáveis, mas não dolorosos. Dentre os sintomas positivos, que são dolorosos, estão parestesia/disestesia (formigamentos, agulhadas), hiperalgesia (resposta exagerada a um estímulo doloroso) e alodinia (dor sentida por um estímulo que normalmente não causaria dor)^{1,2}. Algumas pessoas acometidas pela dor neuropática descreveram a dor sentida como “cruel” e “exaustiva”².

Como alguns sintomas são gerados dos mecanismos de canais iônicos¹, os medicamentos utilizados para o seu tratamento são fármacos com propriedades de bloqueio de canais iônicos (antidepressivos, anticonvulsivantes, analgésicos, opióides, entre outros). No entanto, esses medicamentos podem não aliviar a dor consideravelmente e possuem vários

efeitos colaterais⁷, sendo, além disso, altamente dispendioso para o sistema de saúde⁸.

Apesar do progresso no tratamento da dor neuropática, ainda não se conhece em detalhes como ocorre a progressão da dor. Alguns alvos têm sido estudados, dentre eles alvos moleculares envolvidos nos mecanismos de canais iônicos que, ao serem bloqueados, apresentam utilidade no tratamento da dor neuropática⁹. Um dos alvos desses estudos têm sido o receptor sigma-1^{10,11}.

Quando descobertos, os receptores sigma foram classificados como uma classe de receptores opióides¹². Posteriormente, foi esclarecido que são uma classe particular de proteínas¹³, tendo sido identificados dois tipos de receptores sigma: o receptor sigma-1 e o receptor sigma-2¹⁴. O receptor sigma-2 ainda não foi clonado e caracterizado, porém estudos mostraram que ele pode ser utilizado em tratamentos de câncer^{15,16}.

O receptor sigma-1 já foi clonado e identificado como uma proteína de 223 aminoácidos, se encontrando na interface existente entre o retículo endoplasmático e a mitocôndria, chamada de MAM (do inglês *Mitochondrion-associated ER membrane*) atuando no regulamento da sinalização do íon Ca^{2+} entre o retículo endoplasmático e a mitocôndria^{17,18}. Ele pode, devido à estimulação por ligantes ou *stress* celular, se deslocar para outras áreas da célula, possibilitando com que regule de maneira suplementar uma ampla variedade de canais iônicos, receptores e proteínas quinases^{10-11,19-20}.

Por isso, devido à sua translocação para outras partes da célula e ampla atuação, encontrou-se em estudos variados que receptores sigma-1 podem ser alvos para o tratamento de várias doenças do sistema nervoso central como Alzheimer, esquizofrenia, depressão, câncer e inclusive a dor neuropática¹⁰. Para o tratamento da dor neuropática, a função do receptor sigma-1 pode ser modulada e bloqueada por antagonistas, afetando a habilidade do receptor sigma-1 de interagir com diferentes proteínas²¹, sendo capaz de reduzir a dor.

De acordo com a literatura²², para que compostos atuem como antagonistas do receptor Sigma-1 devem ser formados por um centro contendo amina e com duas regiões hidrofóbicas a distâncias determinadas de 6 a 10 Å e 2,5 a 3,9 Å, conforme mostrado na Figura 1.

Figura 1. Estrutura geral de um composto antagonista do receptor Sigma-1.

Diversos compostos têm sido sintetizados para que possam atuar como antagonistas do receptor sigma-1. Dentre eles, tem-se dois que podem atuar no tratamento da dor neuropática, que são o S1RA (E-52862)²³, composto derivado de uma classe de 1-arilpirazóis que se encontra em fase de testes de medicamentos em seres humanos²⁴ e o BD1047^{25,26}. Suas estruturas moleculares são mostradas na Figura 2.

Figura 2. Estruturas químicas de compostos antagonistas do receptor Sigma-1 que podem atuar no tratamento da dor neuropática.

S1RA

BD1047

Uma série de compostos, tendo como estrutura básica a pirimidina, conforme mostrado na Figura 3, foi sintetizada e apresentou potencial aplicação no tratamento da dor neuropática²⁷.

Figura 3. Estrutura molecular do composto de referência formada pelo esqueleto de pirimidina.

1.1 Métodos Computacionais

Diversos estudos têm sido realizados para o desenvolvimento de fármacos. Dentre eles, pode-se citar estudos de QSAR (do inglês *Quantitative Structure-Activity Relationships*)²⁸⁻³¹, o qual consistem em relacionar a estrutura química de um conjunto de compostos químicos que tenham um esqueleto comum entre si com suas respectivas atividades biológicas, ao fazer uso de métodos computacionais³²⁻³³. Com isso, pode-se prever o mecanismo de ação dos compostos através da geração de modelos matemáticos³⁴⁻³⁶. Esses modelos podem ser construídos e gerados através da seleção de descritores, que são parâmetros quantitativos característicos de cada molécula e fornecem informações sobre suas estruturas.

Os descritores podem ser obtidos através de métodos computacionais, sendo que o método quântico DFT (do inglês *Density Function Theory*) têm sido muito utilizado para tal fim, pois têm se mostrado eficaz por apresentar um custo computacional baixo e resultados satisfatórios. O método de DFT se baseia na resolução da equação de Schrödinger para elétrons de sistemas moleculares em função da posição do núcleo, no qual o estado fundamental de energia de um átomo ou molécula é expresso em termos de sua densidade eletrônica. Ele possibilita realizar o cálculo de propriedades de átomos e moléculas, que são os descritores³⁷ a serem utilizados em estudos QSAR, como eletronegatividade, maciez e dureza, potencial químico, dentre outras³⁸⁻

Os descritores podem ser divididos em várias classes, podendo-se citar os eletrônicos, topológicos, geométricos e constitucionais, por exemplo. Eles possibilitam obter informações sobre a molécula como tamanho e forma, distribuição de cargas, interações intramoleculares, distribuições de energia, dentre muitas outras, tornando possível adquirir informações necessárias para o entendimento da relação entre a estrutura molecular com sua atividade⁴³. Deve-se ter cautela na seleção dos descritores, pois são selecionados dentre uma grande quantidade de variáveis e devem conter as informações mais relevantes para a construção dos modelos QSAR⁴⁴.

Algumas técnicas devem ser empregadas na seleção dos descritores para que possam ser definidos os melhores modelos QSAR. Dentre essas técnicas, podem ser utilizados para a seleção dos descritores a matriz de correlação e o algoritmo genético.

A matriz de correlação consiste em correlacionar as variáveis entre si através do coeficiente de correlação de Pearson (R_{ij}), que se dá através da equação (1), no qual X_i e X_j correspondem aos valores das variáveis e n corresponde ao número de amostras.

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

Para os estudos de QSAR, relacionam-se os valores obtidos de cada descritor com o valor da atividade biológica correspondente. Valores muito baixos de R_{ij} indicam uma correlação muito baixa e valores muito altos indicam uma proximidade muito grande, sendo considerados repetitivos por serem extremamente correlacionados, portanto baixos valores e altos valores de R_{ij} não são desejáveis. No entanto, a faixa de valores a serem considerados deve ser definida de acordo com o estudo a ser realizado. Em estudos de QSAR recomenda-se valores abaixo de 0,6.⁴⁵⁻⁴⁶

A técnica de algoritmo genético identifica e classifica os compostos de acordo com suas características similares e torna possível a seleção dos descritores moleculares importantes para a construção do modelo.

Baseado nos conceitos de evolução, tais como mutação, seleção e cruzamento, são selecionados os melhores indivíduos em uma população dentro de sucessivas gerações. Inicialmente, estabelecem-se as condições a serem alcançadas e são feitas todas as combinações possíveis dentre os indivíduos de forma aleatória. Após sucessivas gerações, vão sendo selecionados os indivíduos que apresentam as melhores características. Então, tem-se o cruzamento e/ou mutação dos indivíduos, até que uma nova população seja gerada. O processo de geração é repetido até que seja alcançada uma solução que satisfaça os critérios estabelecidos³⁵. O esquema geral do funcionamento do algoritmo genético é mostrado na Figura 4.

Figura 4. Esquema geral do funcionamento da técnica de algoritmo genético.

Para estudos QSAR, têm-se os valores de atividade biológica medidos experimentalmente como sendo os valores que se desejam alcançar. A população formada pelos indivíduos são os descritores, de tal forma que são feitas todas as combinações possíveis dentre eles sendo selecionados as combinações que apresentarem os melhores valores de q^2 (coeficiente de correlação da validação cruzada) e R (coeficiente de correlação) após sucessivas gerações, mutação e cruzamento.⁴⁷⁻⁵⁰

Para a geração de modelos QSAR, podem ser utilizados os métodos de Mínimos Quadrados Parciais ou PLS (do inglês *Partial Least Squares*), que torna possível a interpretação dos dados ao desenvolver uma expressão matemática que correlacione linearmente as variáveis selecionadas⁴¹ e Redes

Neurais Artificiais ou ANN (do inglês *Artificial Neural Network*), baseado na arquitetura de uma rede de neurônios que permite a construção do modelo correlacionando as variáveis de forma não linear⁵¹⁻⁵².

1.1.1 Geração do Modelo QSAR - Método PLS

O método de PLS relaciona os valores de duas matrizes X e Y através de regressões lineares múltiplas. Para os estudos QSAR, tem-se que a matriz X corresponde aos valores de atividade biológica preditos e a matriz Y corresponde aos valores dos descritores selecionados. As matrizes X e Y podem ser descritas pelas equações (2) e (3), nas quais T e U representam as matrizes de escores, P' e Q' as matrizes de pesos, E e F as matrizes dos resíduos, t_a e u_a são os elementos das matrizes de escores T e U e p'_a e q'_a são os elementos das matrizes de pesos P' e Q' das matrizes X e Y , respectivamente.

$$X = TP' + E = \sum t_a p'_a + E \quad (2)$$

$$Y = UQ' + F = \sum u_a q'_a + F \quad (3)$$

Para ser considerado um bom modelo PLS, deve-se definir o número ótimo de variáveis latentes ao analisar os valores de coeficiente de determinação (r^2) e coeficiente de validação cruzada (q^2). São considerados valores aceitáveis em análises de PLS de r^2 e q^2 como sendo maiores do que 0,6 e 0,5, respectivamente. Os coeficientes r^2 e q^2 são calculados através das equações (4) e (5), nas quais y_{iexp} indica 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 , y_{med} representa o valor experimental médio de Y e $\hat{y}_{ical}$ representa o valor calculado de Y da i -ésima amostra utilizando o modelo construído com todas as amostras.⁵³⁻⁵⁴

$$q^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ival})^2}{\sum (y_{iexp} - y_{med})^2} \quad (4)$$

$$r^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - y_{med})^2} \quad (5)$$

Devem-se analisar também os parâmetros estatísticos que mostram os valores de erros de predição, que são eles a soma dos quadrados dos erros de predição (do inglês PRESS - *Prediction Residues Error Square Sum*), erro padrão de validação (do inglês SEV - *Standard Error of Validation*), erro padrão de calibração (do inglês SEC - *Standard Error of Calibration*), calculados pelas equações (6), (7) e (8), respectivamente. Para um bom modelo PLS, os valores de erros estatísticos devem ser os menores possíveis.⁵⁵

$$PRESS = \sum (y_{iexp} - \hat{y}_{ival})^2 \quad (6)$$

$$SEV = \left[\frac{PRESS}{I} \right]^{\frac{1}{2}} \quad (7)$$

$$SEC = \left[\frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{I - a - 1} \right]^{\frac{1}{2}} \quad (8)$$

Tem-se também o Teste F, que fornece a razão entre a variabilidade explicada e não explicada pelo modelo, devendo ter o maior valor possível, pois assim pode indicar que é bom modelo por ser capaz de explicar suas variações.⁵⁶

$$Teste F = 1 - \frac{\sum (y_{iexp} - y_{med})^2 - (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - \hat{y}_{ical})^2} \frac{(I - 1)}{VL} - 1 \quad (9)$$

Testes adicionais de validação também podem ser aplicados ao modelo PLS para verificar sua robustez. Tem-se a detecção de *outliers*, que identifica a

presença de amostras atípicas, teste *leave-N-out*, que avalia a robustez do modelo e randomização, que identifica correlações ao acaso entre a variável dependente e os descritores.

1.1.2 Geração do Modelo QSAR – Método ANN

Baseado no funcionamento de neurônios biológicos, que processam informações e adquirem conhecimento, o método ANN tem sido muito utilizado por ser capaz de resolver problemas complexos. Esse método pode ser utilizado em QSAR para a predição da atividade biológica de forma não linear, que é conferida de acordo com a função de transferência utilizada.

Dentre as funções de transferência a serem utilizadas, tem-se a *tansig* (função de transferência tangente hiperbólica sigmoidal) e a *logsig* (função de transferência logarítmica sigmoidal). Tem-se também uma função de transferência linear que pode ser utilizada, a *purelin* (função de transferência linear).

A rede mais comumente utilizada em estudos de QSAR é a MLP-ANN (do inglês *Multilayer Perceptron Artificial Neural Network*), que é do tipo perceptron de multicamadas. Ela é constituída pelas camadas de entrada, camadas intermediárias e camada de saída, nas quais as informações passam pelas camadas de entrada até atingirem a camada de saída. Para estudos de QSAR, a camada de entrada é tida como os descritores e a camada de saída corresponde ao valor da atividade biológica predito, como mostrado na Figura 5.

Figura 5. Esquema geral do funcionamento de uma MLP-ANN.

Para avaliar o desempenho de uma ANN, pode-se analisar os valores dos coeficientes estatísticos obtidos, tais como o coeficiente de determinação (r^2), erro quadrático médio (EQM) e desvio médio absoluto (DMA).⁵⁷⁻⁵⁹

Em virtude de a dor neuropática atingir uma alta porcentagem da população global, do impacto que possui na qualidade de vida das pessoas afetadas por ela, da falta de medicamentos eficazes e dos altos custos para o seu tratamento, dá-se a importância de realizar estudos para encontrar alvos apropriados para o desenvolvimento de fármacos específicos para o seu tratamento. Com os métodos apresentados, é possível realizar estudos QSAR, os quais possibilitam adquirir uma compreensão maior de aspectos relevantes da interação de compostos antagonistas do receptor sigma-1, de forma que possam ser úteis na produção de alvos específicos para o tratamento da dor neuropática.

2 OBJETIVOS

O presente trabalho visa estudar as relações quantitativas entre a estrutura química e a atividade biológica dos antagonistas do receptor sigma-1, consistindo nos seguintes objetivos:

- Realizar a otimização da geometria dos compostos em seu estados neutro, reduzido e oxidado através do método de cálculo DFT para obtenção dos descritores;
- Selecionar os descritores moleculares e os descritores eletrônicos a partir das técnicas de matriz de correlação e algoritmo genético;
- Geração e validação de modelos QSAR através das técnicas PLS e Redes Neurais Artificiais.

3 METODOLOGIA

Selecionou-se da literatura um conjunto de compostos para que pudessem ser realizados estudos de QSAR. Para que se tenha um bom estudo de QSAR, a faixa de variação de atividade biológica deve ser de pelo menos duas ordens de magnitude em unidades de concentração molar e estes valores são transformados na forma logarítmica para esse cálculo⁴⁵. Para o conjunto de compostos selecionado, a atividade biológica se deu em termos de K_i (constante de dissociação do inibidor).

Para que ocorresse a seleção de um grupo de compostos, foi necessário verificar a faixa de variação de atividade biológica e a distribuição que a mesma apresenta, verificação essa que se dá em três partes, que são: primeiro, obter da literatura os valores de atividade biológica determinados experimentalmente (K_i); segundo, calculá-los em sua forma logarítmica em $-\log K_i$ (representado por pK_i) e terceiro, esboçar uma curva de pK_i versus pK_i .

Após realizar essa verificação em vários grupos de compostos encontrados na literatura, selecionou-se um grupo formado por uma série de pirimidinas antagonistas do receptor Sigma-1²⁷. Foram realizados os cálculos dos valores de pK_i e esboço da curva de pK_i versus pK_i , conforme mostrados na Figura 6.

Figura 6. Distribuição dos valores de pK_i para a série de pirimidinas, antagonistas do receptor sigma-1.

Através desses cálculos, pôde-se obter como resultado o valor de aproximadamente 3,30 para a faixa de variação de atividade biológica para esse grupo de compostos. Esse resultado se mostrou satisfatório para que ocorresse a realização de estudos de QSAR com essa classe de compostos por essa faixa de variação encontrada satisfazer a condição de pelo menos duas ordens de magnitude, conforme explicado.

O grupo escolhido tem um total de 54 compostos, que foram selecionados de uma série de moléculas que atuam como antagonistas do receptor Sigma-1 para o tratamento da dor neuropática. A estrutura molecular de referência desse conjunto é formada pelo esqueleto de pirimidina através de mudanças dos radicais R_1 , R_2 , R_3 , R_4 e R_5 conforme se encontra na Figura 3²⁷. Assim, ao efetuar mudanças nesses radicais, têm-se os compostos com as estruturas apresentadas no Apêndice A e também nas figuras 7 a 16.

Para atender às condições do modelo farmacofórico e assim ter compostos que possam atuar como antagonistas do receptor Sigma-1, no composto 1 tem-se que região hidrofóbica primária corresponde a um radical fenil e na região hidrofóbica secundária tem-se o grupo morfolina (Figura 7).

Figura 7. Estruturas moleculares dos compostos 1 ao 6.

Nota-se que nos compostos 1 a 6 (Figura 7) tem-se uma mudança na cadeia de carbonos que liga o esqueleto de pirimidina e a morfolina. O composto 1 se mostrou ativo com três carbonos na cadeia. Com o aumento do número de carbonos na cadeia, nos compostos 2 e 3, observa-se uma diminuição no valor da atividade biológica. O composto 4 apresenta uma insaturação com relação ao composto 1, tendo também uma diminuição no valor de pK_i . O composto 5 tem um átomo de carbono substituído por um átomo de oxigênio em relação ao composto 3, mantendo os valores da atividade bem próximos. Tem-se uma hidroxila como ligante no composto 6 em relação ao composto 1, podendo observar um decaimento significativo no valor da atividade. Nota-se que a distância e as características da cadeia de carbonos interfere significativamente no valor da atividade biológica e o melhor valor de atividade biológica se deu com 3 carbonos na cadeia, sem insaturações, ligantes ou heteroátomos. Portanto, o número da cadeia de carbonos se manteve em três carbonos para o restante dos compostos do conjunto.

Tem-se nos compostos 7 ao 20 (Figuras 8 a 11) substituições de alguns grupos nas posições R_1 e R_2 .

Figura 8. Estruturas moleculares dos compostos 7 ao 10.

Nos compostos 7 ao 10 (Figura 8), tem-se a substituição do grupo morfolina pelos grupos piperidina, 4-metilpiperidina, 3,5-dimetilpiperidina e 2,2,6,6-tetrametilpiperidina, respectivamente. O composto 7 apresentou um maior valor de atividade biológica em relação ao composto 1, que continha a morfolina na região hidrofóbica secundária. Esse aumento se deu provavelmente devido à maior hidrofobicidade da piperidina.

Nos compostos 8, 9 e 10, tem-se a adição de grupos metil na piperidina, mostrando um decaimento no valor da atividade, provavelmente devido ao impedimento estérico, que deve afetar a interação entre o ligante e o receptor.

Figura 9. Estruturas moleculares dos compostos 11 ao 16.

Tem-se a substituição do átomo de oxigênio do composto 1 da morfolina por uma hidroxila e uma carbonila nos compostos 11 e 12 (Figura 9), respectivamente, podendo notar uma diminuição significativa da atividade. Isso mostrou que os grupos polares não foram favoráveis ao se ligar com o receptor Sigma-1 nessa posição. O composto 13 tem a substituição do átomo de oxigênio do composto 1 do grupo morfolina por um átomo de nitrogênio, sendo um derivado de piperazina, e mostrou uma baixa afinidade com o receptor. Tem-se a substituição do hidrogênio na piperazina por pequenos grupos

alquila, que resultaram em compostos mais potentes, os compostos 14, 15 e 16. Isso se deu devido ao aumento da hidrofobicidade na região secundária.

Figura 10. Estrutura molecular do composto 17.

O composto 17 (Figura 10), teve a morfolina do composto 1 substituída por uma pirrolidina, mostrando-se também mais potente.

Figura 11. Estruturas moleculares dos compostos 19 a 20.

Os compostos 18, 19 e 20 (Figura 11), têm aminas de cadeia aberta no lugar da morfolina e apresentam valores da atividade biológica próximos.

Com base nas substituições realizadas dos compostos 7 ao 17 (Figuras 8 a 10), os grupos que apresentaram maior potência para a região hidrofóbica secundária ao atuarem como antagonistas do receptor sigma-1 foram a piperidina, no composto 7, e a pirrolidina, no composto 17. Esses substituintes se mostraram presentes nas posições R₁ e R₂ nos compostos seguintes, sendo que nos compostos restantes têm-se substituições nas posições R₃, R₄ e R₅.

Figura 12. Estruturas moleculares dos compostos 21 a 27 e 31 a 37.

Nos compostos 21 a 27 e 31 a 37 (Figura 12), tem-se a substituição do radical metil (que se encontra na posição R_3) por outros substituintes. As potências dos compostos 21 e 22 se mostraram menores em relação ao composto 7 e da mesma forma ocorreu com os compostos 31 e 32 em relação ao composto 17 (Figura 11). Substitui-se o radical metil pelo radical isopropil nos compostos 23 e 33, pelo ciclopropil nos compostos 26 e 36 e pelo fenil nos compostos 27 e 37, tendo um decréscimo significativo na atividade biológica. Esse decréscimo na substituição na posição R_3 se deu provavelmente devido ao tamanho dos substituintes, por causar impedimento estérico.

Com base nas modificações na posição R_3 , tem-se que os substituintes apresentaram afinidade com o receptor sigma-1 na seguinte ordem: metoxil> metil> trifluorometil > etil> isopropil> n-propil> ciclopropil> fenil.

Figura 13. Estruturas moleculares dos compostos 28 a 30 e 38 a 40.

Na Figura 13, têm-se os compostos 28 a 30 e 38 a 40. Ao se ter substituintes pequenos na posição R_4 nos compostos 28 e 38, tem-se um bom valor de atividade biológica. Halogênios nessa posição, com o flúor nos compostos 29 e 39 e cloro nos compostos 30 e 40, mostraram um valor elevado de atividade biológica, sendo os maiores valores para as substituições feitas até então. O composto 30 com um átomo de cloro na posição R_4 e piperidina nas posições R_1 e R_2 , mostrou maior valor de atividade biológica em relação ao composto 40, que possui um anel de pirrolidina na posição R_1 e R_2 .

Figura 14. Estruturas moleculares dos compostos 41 a 43.

Os compostos 41, 42 e 43 (Figura 14), apresentam anéis com 5 e 6 carbonos ligados nas posições R_3 e R_4 . O anel de cinco carbonos com a piperidina nas posições R_1 e R_2 , composto 41, mostrou maior potência em relação ao composto 42, que tem a pirrolidina nas posições R_1 e R_2 . O composto 43, que tem o anel de 6 carbonos ligado à pirimidina, apresentou uma valor muito baixo atividade biológica, sendo o menor valor desse conjunto de compostos. Isso ocorreu provavelmente devido a um maior impedimento estérico, dificultando a ligação com o receptor.

O composto 30 (Figura 13) com o cloro na posição R_3 apresentou alto valor de atividade biológica, portanto manteve-se o átomo de cloro nessa posição nos compostos 44 a 54 (Figuras 15 e 16).

Figura 15. Estruturas moleculares dos compostos 44 ao 50.

Nos compostos 44 a 50 (Figura 15), têm-se ligantes no radical fenil, sendo que com a presença de um metil, que é um elétron doador (composto 44) ou metoxila e trifluorometila (compostos 45 e 46), que são elétrons receptores, tem-se aumento da atividade biológica. Nos compostos 47 a 50 têm-se halogênios ligados ao fenil, apresentando alta afinidade para o receptor Sigma-1 com altos valores de atividade biológica, sendo que o composto que apresentou o maior valor de atividade biológica dentre todos os compostos foi o composto 48.

Figura 16. Estruturas moleculares dos compostos 51 a 54.

Nos compostos 51 a 54 (Figura 16), tem-se a substituição do radical fenil que se encontra na posição R_5 por um naftaleno, um átomo de hidrogênio, um metil e um ciclopropil, respectivamente. Nota-se a queda dos valores de atividade biológica nesses compostos, sugerindo a importância do anel aromático nessa posição e concluindo então as variações realizadas no conjunto de compostos nas posições R_1 , R_2 , R_3 , R_4 e R_5 .

Então, com o grupo de compostos selecionado da literatura e após ser verificado que esse grupo possui condições satisfatórias para que sejam realizados os estudos em QSAR, pôde-se realizar a construção das estruturas dos antagonistas do receptor Sigma-1 utilizando o software Gaussian 09⁶⁰.

O método de cálculo DFT com o funcional híbrido B3LYP⁶¹ e função de base 6-31g(d,p) foi utilizado para otimizar a geometria dos compostos em seus estados neutros e realizar os cálculos de frequências vibracionais através do Software Gaussian 09. Com as estruturas das moléculas otimizadas, foram obtidos os descritores moleculares através do software Dragon⁶².

Também foram realizados cálculos com as moléculas em seus estados reduzidos e oxidados para a obtenção de descritores eletrônicos, que foram

também extraídos dos cálculos de otimização de geometria e de frequências vibracionais a partir do Software Gaussian 09. Como descritores eletrônicos, foram utilizados os valores de energia da molécula (E), energia no ponto zero (EPZ), polarizabilidade isotrópica, momento dipolo final (μ) e energias dos orbitais de fronteira, HOMO e LUMO, todos no estado neutro da molécula. Através dos cálculos de otimização das moléculas em seus estados reduzidos e oxidados, foram calculados os seguintes descritores eletrônicos: potencial de ionização (PI), afinidade eletrônica (AE), eletronegatividade (χ), eletrofilicidade (ω), dureza (η) e maciez (S), de acordo com as fórmulas (10) a (15), respectivamente.⁵⁰

$$PI = (E_{red} + (EPZ_{red} * 0,9806)) - (E + (EPZ * 0,9806)) \quad (10)$$

no qual E_{red} e EPZ_{red} correspondem à energia da molécula e energia no ponto zero em seu estado reduzido, respectivamente.

$$AE = (E + (EPZ * 0,9806)) - (E_{ox} + (EPZ_{ox} * 0,9806)) \quad (11)$$

no qual E_{ox} e EPZ_{ox} correspondem à energia da molécula e energia no ponto zero em seu estado oxidado, respectivamente;

$$\chi = \frac{(PI + AE)}{2} \quad (12)$$

$$\omega = \frac{\mu^2}{2\eta} \quad (13)$$

$$\eta = \frac{(PI - AE)}{2} \quad (14)$$

$$S = \frac{1}{\eta} \quad (15)$$

Após isso, os estudos quimiométricos das relações estrutura-atividade ocorreram através de duas etapas: seleção de variáveis e geração dos modelos QSAR.

3.1 Seleção de variáveis

Para que ocorresse a seleção das variáveis, que são os descritores, primeiramente utilizou-se a técnica de matriz de correlação utilizando o Excel, de modo que foi calculado o coeficiente de correlação de Pearson R, representado pelo grau de relação linear entre os valores de cada descritor e os valores de pK_i . Os descritores que tiveram o valor de R abaixo de 0,3 em módulo foram descartados por terem uma baixa associação linear entre os valores do descritor e do pK_i .

Após isso, a técnica de algoritmo genético foi realizada com o uso do software BuildQSAR 2.1.0.0⁶³ no qual os descritores selecionados com a técnica de matriz de correlação foram submetidos a uma nova seleção ao gerar diferentes modelos conforme se variava o número de descritores, número de gerações e o critério de corte para a correlação entre os descritores.

Ao concluir a análise da técnica de algoritmo genético, foram então selecionados os descritores a serem utilizados nas análises de PLS e ANN, sendo que foi selecionado o melhor modelo gerado pela técnica de algoritmo genético ao analisar os valores de r (coeficiente de correlação) e de q^2 (coeficiente de correlação da validação cruzada) e dos parâmetros estatísticos s (desvio padrão) e Teste F.

3.2 Geração dos modelos QSAR

Após a seleção dos descritores, iniciou-se a etapa de geração dos modelos. Para geração do modelo através da técnica de PLS, foi utilizado o software Pirouette⁶⁴. Para tal, dividiu-se o conjunto de compostos em um conjunto teste e um conjunto treinamento. O conjunto teste foi formado por cerca de 20% de compostos para cada faixa de variação de 0,5 unidades de pK_i , tendo um total de 11 compostos, conforme destacados na Tabela 2. Para o método de ANN é necessário ter um conjunto validação durante o treinamento

de uma MLP-ANN para impedir o superajuste da rede, que foi formado por 5 compostos. Portanto, esses compostos não entraram no conjunto treinamento para ambos os métodos, para termos de comparação entre os métodos. Assim sendo, o conjunto treinamento foi formado por 38 compostos.

Para que fosse definido o melhor modelo de PLS, foi realizada uma análise dos parâmetros estatísticos SEV, SEC, PRESS, q^2 e r através do software Pirouette. Também foram feitas análises de detecção de outliers para que se caso tivessem amostras com comportamento incomum, pudessem ser identificadas.

Posteriormente, foram feitas validações dos modelos através dos testes *leave-N-out* e randomização da atividade biológica, através do software QSAR Modeling⁶⁵.

Para o teste *leave-N-out*, excluiu-se N compostos do conjunto treinamento, sendo que para cada N repetiu-se as análises por 30 vezes. Definiu-se o valor de N de acordo com o número de compostos presentes no treinamento, correspondendo a cerca de 20% do conjunto.

Foram realizados 50 testes de randomização dos valores da atividade biológica para que pudesse se certificar a obtenção dos modelos a partir de correlações ao acaso. Foram feitos testes de *leave-one-out* para cada randomização, que tiveram os valores de r^2 e q^2 analisados.

Para geração de modelos de ANN, foi utilizado o software MatLab R2011a⁶⁶ utilizando como modelo de ANN a do tipo Perceptron de Multicamadas (do inglês: MLP-ANN – Multilayer Perceptron Artificial Neural Network), a qual tem uma camada de entrada, representada pelos descritores, uma camada intermediária, que é definida de acordo com o número de neurônios que se deseja ter, e uma camada de saída, representada pelos valores preditos de pK_i . Utilizou-se o algoritmo de retropropagação e os valores dos descritores foram escalados entre 0,1 e 0,9, de acordo com a equação (16), para que pudesse ocorrer a normalização de valores máximos e mínimos. Na equação (16) tem-se que i e i' correspondem aos valores não normalizados e normalizados para as entradas de ANN, respectivamente, e $i_{\min}$ e $i_{\max}$ são os valores mínimos e máximos para essas entradas.

$$i' = 0,8 \frac{i - i_{min}}{i_{max} - i_{min}} + 0,1 \quad (16)$$

Realizou-se a predição dos valores de pK_i para o conjunto de compostos ao se combinar de duas a duas das três funções de transferência, resultando nas seguintes combinações: *tansig-tansig*, *tansig-logsig*, *tansig-purelin*, *logsig-logsig*, *logsig-tansig*, *logsig-purelin*, *purelin-purelin*, *purelin-tansig* e *purelin-logsig*. A primeira função da combinação atua na camada intermediária e a segunda função atua na camada de saída durante o treinamento da MLP-ANN. Para cada combinação de funções, variou-se o número de neurônios na camada intermediária de 1 a 20.

4 RESULTADOS E DISCUSSÃO

4.1 Seleção de variáveis

Os descritores moleculares foram obtidos a partir do Software Dragon, que gerou para esse conjunto de compostos um total de 715 descritores moleculares. Ao incluir os descritores eletrônicos que foram obtidos através dos cálculos de otimização de geometria e frequências vibracionais das moléculas, tais como afinidade eletrônica, momento de dipolo, polarizabilidade isotrópica, dentre outros, conforme já mencionados, atingindo então um total de 736 descritores.

Após isso, pode-se dar início à etapa de seleção de variáveis, na qual foram utilizados os métodos de matriz de correlação e algoritmo genético para selecionar os descritores a serem utilizados para a geração dos modelos de QSAR.

Primeiramente realizou-se uma filtragem no número inicial de descritores, através da técnica de matriz de correlação, no qual os descritores que apresentaram valores de coeficiente de correlação de Pearson R menores do que 0,3 foram excluídos. Ao realizar essa análise, restou um total de 104 descritores.

Através do método de algoritmo genético utilizando o Software BuildQSAR, ao variar o número de gerações de 500 até 10000 e número de descritores de 4 a 9, com 5 modelos por geração, foi possível obter 390 modelos. Dentre esses modelos, foram pré-selecionados os que tinham os maiores valores de r e q^2 , sendo que foram considerados 108 modelos com os melhores valores.

Esses 108 modelos tiveram os seus descritores submetidos a análises prévias de PLS. Essas análises prévias consistiram em gerar modelos PLS com todos os compostos do conjunto para cada modelo obtido pela técnica de algoritmo genético, de forma que foram analisados os valores de q^2 , r , SEC, SEV e PRESS. Assim, foi possível selecionar os descritores que apresentaram nessas análises prévias parâmetros estatísticos satisfatórios a fim de gerar os modelos PLS e ANN.

Então, o conjunto de descritores selecionados é composto pelos descritores nCt (número total de carbonos terciários), BELm1 (menor autovalor

número 1 da matriz de Burden ponderado pela massa atômica), J3D (índice de distância de conectividade de Balaban), CIC_0 (conteúdo de informações complementares de simetria de vizinhança de ordem 0), Mor12e (sinal 12 dos descritores 3D-MoRSE ponderado pela eletronegatividade atômica de Sanderson), T(N..O) (soma das distâncias topológicas entre nitrogênio e oxigênio), E_{LUMO+4} (energia do orbital de fronteira LUMO+4) e μ (momento dipolo). Os parâmetros estatísticos do melhor modelo gerado pela técnica de seleções de variáveis por algoritmo genético se encontram na Tabela 1 e os valores dos descritores selecionados para que se pudessem gerar modelos através das técnicas de PLS e ANN se encontram na Tabela 2.

Tabela 1. Parâmetros estatísticos do modelo gerado pela técnica de algoritmo genético.

Descritores:	r¹	s²	F³	q^{2 4}
nCt, BELm1, J3D, CIC0, Mor12e, T(N..O), E_{LUMO+4} e μ	0,927	0,327	34,212	0,798

¹Coeficiente de correlação. ²Desvio padrão. ³Teste F. ⁴Coeficiente de correlação da validação cruzada.

Tabela 2. Valores dos descritores selecionados para realizar a geração de modelos PLS e ANN (os compostos do conjunto teste encontram-se sublinhados).

Composto	E_{LUMO+4}	μ	nCt	BELm1	J3D	CIC0	Mor12e	T(N..O)
1	1,72	3,99	0	1	2	4	-1	31
2	1,80	3,73	0	1	2	4	-1	34
3	1,63	4,20	0	1	2	4	-2	37
4	0,18	2,85	0	1	1	4	-1	34
5	1,38	2,97	0	1	2	4	-1	52
6	1,46	5,12	0	1	2	4	-1	46
7	2,13	3,38	0	1	2	4	-2	8
8	1,61	3,37	1	1	2	4	-2	8
9	1,82	3,33	2	1	2	4	-2	8
10	1,54	2,99	2	2	2	4	-3	8
11	1,31	4,63	0	1	2	4	-2	34
12	1,47	4,61	0	1	2	4	-1	34
13	1,37	3,24	0	1	2	4	-2	15

14	1,75	3,16	0	1	2	4	-2	15
15	2,04	3,09	0	1	2	4	-1	15
16	1,70	2,96	0	1	2	4	-2	15
17	1,86	3,40	0	1	2	4	-1	8
18	2,29	3,36	0	1	2	4	-1	8
19	2,26	3,24	0	1	2	4	-2	8
20	1,93	3,19	0	1	2	4	-2	8
21	0,24	3,42	0	1	2	4	-2	8
22	1,53	3,48	0	1	2	4	-2	8
23	1,53	3,28	1	1	2	4	-2	8
24	1,21	0,71	0	1	2	4	-2	20
25	1,46	3,39	0	1	2	3	-1	8
26	1,52	3,26	1	1	1	4	-2	8
27	0,09	3,23	0	2	1	4	-2	8
28	1,50	1,95	0	1	2	4	-1	8
29	1,56	2,32	0	1	2	4	-1	8
30	0,83	1,74	0	1	2	4	-1	8
31	1,87	3,34	0	1	2	4	-1	8
32	1,87	3,31	0	1	2	4	-2	8
33	1,87	3,25	1	1	2	4	-2	8
34	1,65	2,30	0	1	2	4	-1	20
35	1,78	3,47	0	1	2	3	-1	8
36	1,87	3,35	1	1	1	4	-2	8
37	0,12	3,31	0	2	1	4	-2	8
38	1,85	1,95	0	1	2	4	-1	8
39	1,87	2,40	0	1	2	3	0	8
40	0,18	1,51	0	1	2	3	0	8
41	1,58	1,92	0	1	1	4	-1	8
42	1,92	2,17	0	1	1	4	-1	8
43	1,99	1,85	0	2	1	4	-2	8
44	0,20	1,42	0	1	2	4	-2	8
45	0,13	2,48	0	1	2	4	-1	32
46	0,02	2,93	0	1	2	3	-1	8

47	0,09	2,68	0	1	2	3	0	8
48	0,01	3,78	0	1	2	3	0	8
49	0,05	2,05	0	1	2	3	-1	8
50	1,36	4,43	0	1	2	3	-1	8
51	0,11	1,47	0	2	1	4	-1	8
52	2,16	1,27	0	1	2	3	-1	8
53	2,18	1,60	0	1	2	3	-1	8
54	2,18	1,72	0	1	2	3	-1	8

4.2 Geração de modelos QSAR

Na etapa de geração de modelos QSAR, foram utilizados os métodos de Mínimos Quadrados Parciais (PLS) e de Redes Neurais Artificiais (ANN).

4.2.1 Mínimos Quadrados Parciais (PLS)

Através dos estudos de PLS, foram propostos modelos formados por 38 compostos no conjunto treinamento, validados externamente com 11 compostos do conjunto teste. Selecionou-se o melhor modelo, formado pelos conjuntos treinamento e teste conforme mostrados na Tabela 2. Esse modelo tornou possível a predição dos valores de pK_i do conjunto de compostos antagonistas do receptor sigma-1 de modo satisfatório. Na Tabela 3 se encontram os parâmetros estatísticos do modelo gerado.

Tabela 3. Variância percentual acumulada, SEV, SEC, PRESS, q^2 e r^2 dos modelos PLS obtidos para o conjunto treinamento.

VL	%Var _{acumulada} ¹	SEV ²	PRESS _{val} ³	SEC ⁴	PRESS _{cal} ⁵	q^2 ⁶	r^2 ⁷
1	24,53	0,469	8,564	0,423	6,628	0,662	0,738
2	33,83	0,398	6,177	0,313	3,525	0,761	0,861
3	43,09	0,363	5,144	0,306	3,277	0,798	0,870
4	59,13	0,363	5,129	0,306	3,181	0,800	0,874
5	71,42	0,367	5,249	0,308	3,128	0,796	0,876
6	81,46	0,363	5,138	0,312	3,116	0,800	0,877
7	91,77	0,362	5,121	0,317	3,113	0,800	0,877
8	100,00	0,362	5,118	0,322	3,113	0,800	0,877

¹ Variância percentual acumulada. ² Erro padrão de validação. ³ Soma dos quadrados dos erros de previsão da validação. ⁴ Erro padrão de calibração. ⁵ Soma dos quadrados dos erros de previsão da calibração. ⁶ Coeficiente de correlação de validação cruzada. ⁷ Coeficiente de determinação múltipla.

De acordo com a literatura⁶⁷⁻⁶⁸, resultados são considerados satisfatórios para modelos que apresentem valores de $q^2 > 0,5$ e $r^2 > 0,6$. Como se nota na Tabela 3, a partir de 6 variáveis latentes não se tem mudanças nos valores de q^2 e r^2 , sendo que os modelos que apresentam os maiores valores para tais variáveis são com 6, 7 e 8 variáveis latentes. Porém, nota-se que para 6 variáveis latentes, o valor de SEC é o menor e o valor de SEV é o segundo menor. Portanto, foi escolhido que para 6 variáveis latentes seria o número ótimo de fatores.

A contribuição dos descritores para o modelo PLS é mostrada na Figura 17. Nota-se que os descritores que apresentam maiores valores de coeficientes são BELm1, nCt e E_{LUMO+4}, sendo portanto os três descritores mais significativos na descrição da atividade biológica do conjunto de compostos em estudo.

Figura 17. Coeficientes dos descritores para o modelo PLS.

Realizou-se a análise de amostras atípicas para garantir que o modelo não contenha compostos com comportamentos anômalos em relação ao conjunto. Essa análise se dá através do gráfico de resíduos de Student *versus* alavancagem, no qual cada amostra é representada por um ponto, como mostrado na Figura 18. A linha vertical representa a alavancagem, sendo que as amostras devem apresentar valores menores que o demarcado por esta linha. As retas horizontais representam a faixa em que se devem encontrar os valores aceitáveis de resíduos de Student.

Figura 18. Detecção de outliers no modelo PLS.

Como pode-se notar na Figura 18, os compostos 5, 24 e 43 apresentaram valores altos de alavancagem, porém com valores de resíduos de Student aceitáveis. O composto 49 se encontra fora dos limites de resíduos de Student, mas dentro dos valores aceitáveis de alavancagem.

Analisou-se as estruturas moleculares desses compostos que apresentaram destaque em relação aos demais. A Figura 19 mostra as estruturas moleculares dos compostos que apresentaram altos valores de alavancagem e a Figura 20 mostra o composto acima do limite de resíduos de Student.

Figura 19. Estruturas moleculares dos compostos que apresentaram altos valores de alavancagem.

Como mostrado na Figura 19, no composto 5 nota-se a presença de um oxigênio a mais na cadeia de carbonos, sendo que esse composto é o único no conjunto treinamento que apresenta tal característica. O composto 24 também é o único que apresenta na posição R_3 o substituinte $O-CH_3$ e, da mesma forma, o composto 43 é o único que apresenta nas posições R_3 e R_4 um anel de 6 carbonos.

Figura 20. Estrutura molecular do composto que apresentou valor fora do limite de resíduos de Student.

O composto 49, como mostrado na Figura 20, tem um átomo de cloro ligado ao anel benzênico na posição para, sendo também o único a apresentar tal ligante no conjunto treinamento.

Pôde-se concluir que as características únicas no conjunto treinamento de tais compostos podem ter contribuído para que os mesmos se tornassem notórios nas análises de *outliers*. Porém, como não foi detectado nenhum composto que possui altos valores de alavancagem e esteja também fora dos

limites de resíduos de Student, nenhum composto foi excluído e deu-se continuidade às análises de validação interna do modelo PLS.

Foram então realizados testes de validação cruzada com a exclusão de N compostos (*leave-N-out*), para verificar a robustez do modelo. Foram feitos também testes de randomização da atividade biológica, para verificar a presença de correlações ao acaso.

Para os testes de *leave-N-out*, foram excluídos de 1 a 10 compostos, sendo feitas 30 repetições para cada valor de N excluído. A Figura 21 mostra os testes de *leave-N-out*.

Figura 21. Testes de *leave-N-out* aplicados ao modelo PLS.

De acordo com a literatura, o modelo é considerado robusto, ou seja, insensível à retirada de amostras, quando atende a duas condições: os valores de q^2 são próximos do valor da retirada de uma amostra (o qual apresenta desvio médio nulo) e quando os desvios de q^2 se encontram na faixa de $\pm 0,05$. De acordo com os dados obtidos, o valor de q^2 com a retirada de uma amostra se manteve em 0,797, com desvio nulo, e o maior valor de desvio médio é de

0,027, que ocorreu com a retirada de 10 amostras. Pode-se então confirmar a robustez e estabilidade do modelo.

Figura 22. Testes de randomização aplicados ao modelo PLS.

Na Figura 22 é possível observar os testes de randomização. De acordo com a literatura, valores de q^2 e r^2 aceitáveis para um modelo randomizado devem ser menores que 0,3 e 0,4, respectivamente. O maior valor obtido para q^2 foi de -0,056 e para r^2 foi de 0,380, sendo, portanto valores aceitáveis. Além disso, os valores do intercepto para $|r|$ versus q^2 deve ser menor que 0,05 e para $|r|$ versus r^2 deve ser menor que 0,3. Tais valores obtidos nos gráficos correspondem a -0,812 e 0,130. Portanto, através dessas análises pôde-se concluir que o modelo não possui correlações ao acaso.

Após a realização da validação interna, feita através dos testes e análises mencionados, foi feita então a validação externa. Através da predição dos valores de pK_i do conjunto teste, obteve-se os valores de $SEV = 0,429$, $PRESS = 2,022$ e $r^2_{pred} = 0,738$. A Figura 23 mostra os valores de pK_i experimental e predito do conjunto teste e a Tabela 4 contém os dados dos resíduos de predição, que correspondem ao erro da predição do conjunto teste. O resíduo relativo de cada composto foi calculado de acordo com a fórmula (17).

$$Resíduo\ relativo = \frac{|Resíduo\ absoluto|}{pk_i\ experimental} \cdot 100 \quad (17)$$

Figura 23. Predição da atividade biológica pelo modelo PLS.

Tabela 4. Resíduos de predição do conjunto teste pelo modelo PLS.

Composto	pK_i experimental	pK_i predito	Resíduo absoluto	Resíduo relativo (%)
1	7,440	7,182	0,259	3,475
6	6,313	6,083	0,230	3,641
8	7,777	7,482	0,295	3,798
9	7,520	7,410	0,110	1,460
12	6,496	7,071	-0,575	8,847
35	7,559	7,397	0,162	2,146
36	6,556	6,420	0,136	2,078
44	8,708	8,441	0,267	3,063
48	9,013	8,449	0,564	6,259
52	6,728	7,724	-0,996	14,806
54	7,403	7,618	-0,215	2,902

Com exceção do composto 52, os compostos do conjunto teste apresentaram resíduos relativos menores do que 10%, o que mostra que o conjunto tem alta capacidade preditiva. O composto 52 apresentou resíduo de 14,806%. Isso pode ser explicado pelo fato de que tal composto é o único que não apresenta um grupo substituinte contendo carbonos no radical R_5 , como

mostrado na Figura 24, e dessa forma apresentou resíduo um pouco maior em relação aos demais compostos do conjunto teste.

Figura 24. Estrutura molecular do composto 52.

Ao realizar os testes de validação interna e externa, pode-se verificar que o modelo PLS se mostrou aceitável por apresentar resultados adequados de acordo com o indicado pela literatura para os parâmetros estatísticos e para os testes de validação. Conforme mostrado no gráfico de resíduos de Student versus alavancagem, não se tem compostos com comportamentos atípicos. Os testes de *leave-N-out* e randomização puderam constatar a robustez do modelo e inexistência de correlações ao acaso. E por fim, através da predição dos valores da atividade biológica utilizando o conjunto teste, o modelo mostrou-se altamente preditivo.

Portanto, o modelo PLS mostrou-se conveniente na predição da atividade biológica do conjunto de compostos antagonistas do Sigma 1-R em estudo de forma linear.

4.2.2 Redes Neurais Artificiais (ANN)

Através dos modelos de ANN, é possível verificar se existe uma relação não linear entre os descritores e a atividade biológica. A rede utilizada para prever os valores de pK_i foi do tipo MLP-ANN.

Para cada uma das variações de funções de transferência e número de neurônios na camada intermediária de MLP-ANN treinadas, foram selecionadas as que apresentaram os melhores valores de r^2 , EQM e DMA, como mostrados no apêndice B. A Figura 24 mostra o desempenho das MLP-ANNs treinadas através do erro quadrático médio (EQM) na predição do pK_i .

Figura 25. Desempenho das MLP-ANNs treinadas.

De acordo com os dados obtidos, pode-se verificar que as combinações de funções de transferência que apresentaram os menores valores de EQM e DMA e os maiores valores de r^2 foram *tansig-tansig*, *purelin-tansig*, *logsig-tansig* e *logsig-purelin*.

Analisou-se os resultados de EQM, DMA e r^2 e também os valores de r^2 para os conjuntos de treinamento, validação e teste, então determinou-se um número ótimo de neurônios na camada intermediária para cada combinação de função de transferência, como mostrado na Tabela 5.

Tabela 5. Melhores modelos de MLP-ANN gerados para cada combinação de função de transferência.

Função de transferência	Arquitetura	EQM	r^2	DMA	$r^2_{\text{treinamento}}$	$r^2_{\text{validação}}$	r^2_{teste}
tansig-tansig	8-11-1	0,010	0,731	0,077	0,734	0,753	0,676
tansig-purelin	8-16-1	0,009	0,742	0,074	0,769	0,733	0,802
tansig-logsig	8-18-1	0,018	0,509	0,093	0,680	0,850	0,712
purelin-tansig	8-15-1	0,009	0,748	0,080	0,759	0,790	0,713
purelin-purelin	8-6-1	0,007	0,800	0,065	0,829	0,836	0,748
purelin-logsig	8-4-1	0,021	0,440	0,105	0,628	0,725	0,631
logsig-tansig	8-4-1	0,009	0,769	0,074	0,752	0,782	0,933
logsig-purelin	8-18-1	0,009	0,753	0,074	0,754	0,712	0,832
logsig-logsig	8-15-1	0,019	0,489	0,099	0,660	0,788	0,660

A MLP-ANN gerada utilizando a combinação de funções de transferência *tansig-tansig* com 11 neurônios na camada intermediária foi escolhida como sendo o modelo que se mostrou mais estável frente às outras combinações de funções de transferência durante o processo de treinamento das MLP-ANN's e por ter valores satisfatórios de r^2 , EQM e DMA.

Tem-se o valor de EQM para o modelo selecionado de 0,010, o valor de r^2 de 0,731 e o valor de DMA de 0,077. Para os valores de r^2 dos conjuntos treinamento, validação e teste, tem-se 0,734, 0,753 e 0,676.

A Figura 26 apresenta os valores experimentais e preditos pelo modelo MLP-ANN selecionado e a Tabela 6 contém os dados dos resíduos de predição do conjunto teste.

Figura 26. Predição da atividade biológica pelo modelo ANN.

Tabela 6. Resíduos de predição do conjunto teste pelo modelo ANN.

Composto	pk _i experimental	pk _i predito	Resíduo absoluto	Resíduo relativo(%)
1	7,440	7,177	0,263	3,531
6	6,313	6,317	-0,004	0,058
8	7,777	7,526	0,252	3,236
9	7,520	7,530	-0,010	0,130
12	6,496	6,886	-0,390	5,996
35	7,559	7,774	-0,215	2,848
36	6,556	6,531	0,025	0,379
44	8,708	7,840	0,868	9,966
48	9,013	8,591	0,422	4,679
52	6,728	7,703	-0,975	14,494
54	7,403	7,695	-0,292	3,945

Os compostos do conjunto teste apresentaram resíduos relativos menores do que 10%, com exceção do composto 52, que apresentou resíduo de aproximadamente 14%, o que ocorreu também no modelo PLS, como já discutido. Esses dados mostram que o modelo MLP-ANN também apresenta alta capacidade preditiva e dessa forma a atividade biológica do conjunto de compostos antagonistas do Sigma-1 R também podem ser descritas de forma não-linear.

4.2.3 Descritores Selecionados

Os descritores que foram selecionados através da técnica de algoritmo genético para a geração de modelos PLS e ANN são mostrados na Tabela 7 em ordem decrescente de significância na descrição da atividade biológica, bem como sua definição e a classe de descritores a que pertencem.

Tabela 7. Descritores selecionados para os modelos QSAR.

Descritor	Definição	Classe
BELm1	Menor autovalor número 1 da matriz de Burden ponderado pela massa atômica	Descritores de Burden (BCUT)
nCt	Número total de carbonos terciários (sp^3)	Grupos funcionais
E_{LUMO+4}	Energia do Orbital LUMO +4	Eletrônicos
Mor12e	Sinal 12 ponderado pela eletronegatividade de Sanderson	Descritores de Morse 3D
J3D	Índice de distância de conectividade de Balaban	Geométricos
μ	Momento Dipolo	Eletrônicos
T(N..O)	Soma das distâncias topológicas entre nitrogênio e oxigênio	Topológicos
CIC₀	Conteúdo de informações complementares (simetria de vizinhança de ordem 0)	Topológicos

- BELm1

O descritor BELm1 pertence à classe de descritores BCUT, os quais são obtidos a partir de autovalores positivos e negativos de uma matriz que contém informações de conectividade e propriedades atômicas, sendo ponderados por propriedades como carga atômica, polarizabilidade, dentre outras. O descritor BELm1 é ponderado pela massa molecular⁶⁹⁻⁷¹.

Pode-se notar que os compostos que apresentaram o valor do descritor BELm1 como sendo 2 (Figura 27) têm um aumento da massa molecular e menores valores de atividade biológica em relação aos outros compostos do conjunto que mostraram o valor para esse descritor igual a 1. Dentre os compostos mostrados na Figura 27, os compostos 27, 37 e 43 tiveram substituintes com maior massa molecular na posição R₃ da estrutura molecular do composto de referência e são os que apresentam os menores valores de pK_i. Portanto, pode-se concluir que essa área tem maior importância para interação molecular com o receptor sigma -1, pois um maior impedimento estérico nessa região acarreta na diminuição da atividade biológica.

- nCt

O descritor nCt indica a presença de carbonos terciários na molécula⁷² com hibridização do tipo sp³. Tem-se a presença de carbonos terciários sp³ para os compostos 8 a 11, 23, 26, 33, 36 e 54.

Os compostos 8 e 9 apresentados na Figura 28, têm a presença de carbonos terciários sp³ nos radicais metil na morfolina e mostraram valores de atividade biológica intermediários por aumentarem a hidrofobicidade na região secundária, conforme proposto pelo modelo farmacofórico. Porém, no composto 10 tem-se uma queda do valor da atividade biológica devido à presença de quatro grupos metila que formam um grande impedimento estérico em volta do átomo de nitrogênio e assim afetam a interação entre o ligante e o receptor.

Figura 28. Compostos com carbonos sp^3 na região hidrofóbica secundária.

Os compostos 23, 33, 26 e 36 mostrados na Figura 29, apresentaram baixos valores de atividade biológica. Pode-se notar que todos eles têm a presença de carbonos terciários na posição R_3 da estrutura molecular do composto de referência. Por essa área ter maior importância na interação com o receptor sigma-1, pode-se concluir que compostos com mais ramos nessa posição se mostram menos ativos, como também mostrado pelo descritor BELm1.

Figura 29. Compostos com carbono sp^3 na posição R_3 .

No composto 54 (Figura 30), tem-se um valor maior da atividade biológica intermediária com a presença de um carbono terciário sp^3 na posição R_5 , sugerindo a importância da região hidrofóbica conforme proposto pelo modelo farmacofórico.

Figura 30. Composto com carbono sp^3 na região hidrofóbica primária.

Composto 54

$pK_i = 7,40$

Portanto, a presença de carbonos terciários sp^3 poderá contribuir para o aumento ou diminuição da atividade biológica de um composto de acordo com a região em que se encontra da molécula. Carbonos terciários sp^3 podem ser desejáveis nas regiões hidrofóbicas por potencializarem a atividade biológica, porém grandes ramificações nas regiões de maior interação molecular com o receptor sigma-1 podem ser indesejáveis por causarem impedimento estérico, impedindo essa interação.

- E_{LUMO+4}

Os orbitais HOMO (do inglês Highest Occupied Molecular Orbital) e LUMO (do inglês Lowest Unoccupied Molecular Orbital) se referem aos orbitais moleculares de fronteira e podem indicar a reatividade química de um composto. Quanto menor a energia de um orbital HOMO, maior será a capacidade da molécula de doar elétrons. Quanto menor a energia de um orbital LUMO, maior será a capacidade da molécula de sofrer ataque nucleofílico e aceitar elétrons⁷³.

No caso do descritor selecionado, E_{LUMO+4} , tem-se os valores de energia do quinto menor orbital desocupado de cada composto. Quanto menor o valor desse descritor, maior a probabilidade do composto em aceitar elétrons.

Ao analisar os dados do descritor E_{LUMO+4} para os compostos antagonistas do receptor Sigma-1, pode-se notar que, os compostos que apresentam maior atividade biológica, com valores de pK_i acima de 8 (Figura 31), com exceção dos compostos 29 e 39 (Figura 32), apresentaram os menores valores de E_{LUMO+4} .

Pode-se notar nesses compostos a energia na região da amina e da região hidrofóbica primária, devido à influência do átomo de cloro, sugerindo

maior interação nessa região. Pode-se notar ainda que, na presença de ligantes apolares ligados ao anel benzênico, nos compostos 44 e 46 tem-se a concentração da energia somente na região da amina, devido à potencialização da região hidrofóbica.

Figura 31. Compostos que apresentaram baixos valores de E_{LUMO+4} .

Os compostos 29 e 39 (Figura 32) mostraram-se com valores altos de atividade biológica, porém apresentaram valores altos para o descritor E_{LUMO+4} . Isso provavelmente se dá devido ao fato de que nesses compostos tem-se a presença de um átomo de flúor na posição R_3 , tendo um deslocamento da energia em relação aos compostos que contém o átomo de cloro na mesma posição tendendo para a região hidrofóbica secundária.

Figura 32. Compostos com baixo valor de pK_i e alto valor de E_{LUMO+4} .

- Mor12e

Os descritores da classe MoRSE (do inglês *3D Molecule Representation of Structures based on Electron diffraction*) fornecem informações tridimensionais da molécula baseadas na difração de elétrons. O descritor Mor12e é ponderado pela eletronegatividade⁷⁴.

Os compostos 39, 40, 47 e 48 (Figura 33) apresentaram valor igual a zero para o descritor Mor12e. Esses compostos estão entre os que apresentaram valores mais altos de atividade biológica.

Figura 33. Compostos que apresentaram valor nulo para o descritor Mor12e.

Esses compostos têm a presença de halogênios, que são altamente eletronegativos, na posição R_3 e nas posições 3 e 4 no anel benzênico. Portanto, o descritor Mor12e mostrou que com a presença de halogênios, têm-se os compostos mais ativos.

- J3D

O descritor J3D é um descritor geométrico que pode caracterizar uma molécula com relação ao seu tamanho. Ele utiliza o índice de conectividade de Balaban (J), que calcula a soma das distâncias médias de conectividade topológicas em um grafo molecular. Como o descritor J3D é um descritor que se dá em três dimensões, então o índice J calcula a distância geométrica ao invés da distância topológica⁷⁵.

Os compostos que apresentaram valor igual a 1 para esse descritor apresentam valores baixos de atividade biológica, que foram os que apresentaram um tamanho maior da cadeia, no caso do composto 4, ou substituintes com maior número de carbonos, como pode-se ver nos outros compostos representados na Figura 34.

Figura 34. Compostos que apresentaram valor igual a 1 para o descritor J3D.

Nos compostos 26, 27, 36, 37, 41 a 43 e 51 têm-se substituintes volumosos na posição R_3 do composto de referência, aumentando a distância geométrica em relação aos outros compostos do conjunto e, como já explicado anteriormente na análise dos descritores BELm1 e nCt, compostos com tal característica implica em menores valores de atividade biológica, não sendo desejáveis para atuar como antagonistas do receptor sigma-1.

- μ

O descritor eletrônico momento dipolo descreve a polaridade de ligação global em uma molécula. Ele pode evidenciar a densidade eletrônica no conjunto de compostos em estudo com a presença de átomos de diferentes eletronegatividades⁷³.

De forma geral, os compostos que apresentaram maiores valores de pK_i tiveram seus valores de momento dipolo menores do que 3 e os menores valores de pK_i apresentaram valores acima de 3 para o momento dipolo. Pode-se destacar que os compostos que apresentaram os maiores valores de atividade biológica foram os que apresentaram a presença dos halogênios flúor e cloro em sua estrutura molecular, que são os átomos mais eletronegativos. Como o flúor é mais eletronegativo que o cloro, então as moléculas que têm o átomo de flúor apresentam maiores valores de momento dipolo em relação às moléculas que contêm átomos de cloro, como nota-se na Figura 35.

Figura 35. Principais compostos do descritor momento dipolo.

- Descritor T(N..O)

O descritor T(N..O) descreve a soma das distâncias topológicas entre átomos de nitrogênio e oxigênio. Quanto maior o valor do descritor T(N..O) para um composto, tem-se uma cadeia carbônica maior ou um maior número de átomos de nitrogênio e/ou oxigênio em sua estrutura molecular⁷⁶.

Para o conjunto de compostos em estudo, tem-se a variação do tamanho da cadeia carbônica para os compostos 1 a 3, como mostrado na Figura 36. Com o aumento de número de carbonos na cadeia carbônica que liga a pirimidina e a morfolina, pode-se notar que se tem uma diminuição na atividade biológica e tem-se que, quanto maior a cadeia carbônica, maior o valor para o descritor T(N..O). Então, para esse caso, quanto maior o valor do descritor, maior será a cadeia carbônica e menor o valor da atividade biológica, portanto moléculas que apresentam altos valores desse descritor não são desejáveis.

Figura 36. Compostos com alteração no tamanho da cadeia de carbonos que liga a pirimidina e a morfolina.

Na Figura 37, têm-se os compostos 4, 5 e 6. O composto 4 apresenta uma insaturação com relação ao composto 2. Com isso, tem-se uma diminuição no valor de pK_i, mas mantém-se o valor do descritor T(N..O) por não haver alteração no tamanho da cadeia. No composto 5, tem-se um átomo de carbono substituindo um átomo de oxigênio em relação ao composto 3, o que manteve os valores da atividade bem próximos (ambos valores baixos), porém aumentou o valor do descritor do composto 5 por se ter a presença de um átomo de oxigênio a mais na molécula. No composto 6 tem-se uma

hidroxila a mais em relação ao composto 1, tendo um decaimento significativo no valor da atividade e também aumento no valor do descritor T(N..O).

Figura 37. Compostos com alterações na cadeia de carbonos.

Nos compostos 11 e 12, conforme mostrados na Figura 38, tem-se a substituição de um átomo de oxigênio do composto 1 da morfolina por uma hidroxila e uma cetona, respectivamente. Com isso, obteve-se uma diminuição significativa da atividade e um pequeno aumento no valor do descritor T(N..O), sugerindo que os grupos polares não foram propícios ao se ligar com o receptor Sigma-1 nessa posição.

Figura 38. Compostos com alteração na morfolina.

Os compostos representados na Figura 39 apresentaram o valor de T(N..O) igual a 15 por terem um átomo de nitrogênio na mesma posição. No composto 13 tem-se a substituição do átomo de oxigênio do composto 1 do grupo morfolina por um átomo de nitrogênio, resultando em um derivado de piperazina, tendo uma diminuição no valor da atividade biológica, que pode se dar devido à propriedade hidrofóbica da piperazina. Nos compostos 14, 15 e

16, substituiu-se o hidrogênio na piperazina com pequenos grupos alquila e ocorreu aumento da propriedade hidrofóbica, resultando em compostos mais potentes.

Figura 39. Compostos que apresentaram valor igual a 15 para o descritor T(N..O).

Os compostos 24 e 34 apresentaram valor do descritor T(N..O) como sendo igual a 20, sendo que tem-se nas posições R₃ uma metoxila, aumentando o valor da atividade biológica. O composto 45 tem uma metoxila ligado ao anel benzênico e, portanto, devido à maior distância entre os átomo de nitrogênio e oxigênio apresentou maior valor do descritor T(N..O) e também mostrou maior valor de atividade biológica (Figura 40).

Figura 40. Compostos com alterações nas posições R₃ e na região hidrofóbica primária.

Pode-se concluir que o descritor T(N..O) é um bom descritor para esse conjunto de compostos ao se analisar os valores obtidos em conjunto com a estrutura molecular do compostos.

- ClC_0

Esse descritor descreve a conectividade e ramificação em uma molécula, podendo estar relacionado à forma molecular e à simetria. Ele representa a diferença entre a complexidade máxima possível de um grafo molecular (no qual cada vértice pertence à uma classe separada de equivalência) e a informação topológica realizada de uma espécie química definida pelo índice de conectividade⁷⁷.

Pode-se notar que todos os compostos que apresentaram valor desse descritor igual a 3 têm a presença de pelo menos um halogênio na molécula, como pode-se ver na Figura 41.

Figura 41. Compostos que apresentaram valor igual a 3 para o descritor ClC_0 .

Portanto, para esse conjunto de compostos esse descritor foi útil para acusar todas as moléculas que têm a presença de halogênios, diferenciando-as do restante do conjunto.

5. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Os métodos propostos se mostraram eficazes para prever a atividade biológica da série antagonistas do receptor sigma-1, pois ao realizar testes de validação, os resultados se mostraram satisfatórios para ambos os métodos, comprovando suas capacidades preditivas.

Os métodos matriz de correlação e algoritmo genético permitiram a seleção de descritores que tiveram valores aceitáveis de r e q^2 . Os descritores selecionados nCt , BE_{LM1} , $J3D$, CIC_0 , $Mor12e$, $T(N..O)$, E_{LUMO+4} e μ se mostraram coerentes na geração de modelos PLS e ANN. Os três descritores que apresentaram maior peso na descrição do modelo, os descritores BE_{LM1} , nCt e E_{LUMO+4} , se mostraram notórios por evidenciar a região de maior importância de interação molecular com o receptor Sigma-1.

Para o modelo PLS, a seleção de 8 descritores se mostrou ideal, sendo que os testes de validação comprovaram sua capacidade preditiva e sua robustez e estabilidade, mostrando também que o modelo não foi obtido a partir de correlações ao acaso.

Para o modelo ANN, obteve-se que a arquitetura 8-11-1, treinada com as funções de transferência *tansig-tansig*, se mostrou a mais adequada para descrever a atividade biológica.

De acordo com os resultados obtidos, conclui-se que a atividade biológica para essa série de pirimidinas antagonistas do receptor Sigma-1 pode ser descrita tanto de forma linear (PLS) quanto de forma não linear (ANN). Portanto, a partir dos modelos de PLS e ANN, é possível direcionar o planejamento de fármacos a serem utilizados no tratamento da dor neuropática.

Visa-se realizar futuramente estudos de *docking* e simulações por dinâmica molecular para verificação do comportamento desse sistema. Outros estudos de QSAR que utilizem outros tipos de descritores também podem ser realizados, como por exemplo, o HQSAR, de tal forma que possam evidenciar novos fatores associados à atividade biológica e complementar os resultados já obtidos nesse estudo.

REFERÊNCIAS

1. NICKEL, F. T.; SEIFERT, F.; LANZ, S.; MAIHÖFNER, C. Mechanisms of neuropathic pain. **European Neuropsychopharmacology**, v. 22, p. 81–91, 2012.
2. GALLUZZI, K. E. Managing neuropathic pain. **The Journal of the American Osteopathic Association**, v. 107, n. 11, p. ES39–ES48, 2007.
3. TREEDE, R.-D.; JENSEN, T. S.; CRUCCU, G.; DOSTROVSKY, J. O.; GRIFFIN, J. W.; HANSSON, P.; HUGHES, R.; NURMIKKO, T.; SERRA, J. Neuropathic Pain Redefinition and a grading system for clinical and research purposes. **Neurology**, v. 70, n. 18, p. 1630-1635, 2008.
4. JENSEN, T. S.; BARON, R.; HAANPÄÄ, M.; KALSO, E.; LOESER, J. D.; RICE, A. S. C.; TREEDE, R.-D. A new definition of neuropathic pain. **Pain**, v. 152, n. 10, p. 2204–2205, 2011.
5. HECKE, O V.; AUSTIN, S.K.; KHAN, R.A.; SMITH, B.H.; TORRANCE, N. Neuropathic pain in the general population: A systematic review of epidemiological studies. **Pain**, v. 155, n. 9, p. 654-662, 2014.
6. TORRANCE N, SMITH BH, BENNETT MI, LEE AJ. The epidemiology of chronic pain of predominantly neuropathic origin. Results from a general population survey. **Pain**, v. 7, n. 4, p. 281- 289, 2006.
7. WOOLF, C. J.; MANNION, R. J. Neuropathic Pain: aetiology, symptoms, mechanisms, and management. **The Lancet**, v. 353, p. 1959-1964, 1999.
8. NIGHTINGALE, S. The neuropathic pain market. **Nature Reviews Drug Discovery**, v. 11, p. 101–102, 2012.
9. STREET, F. B. Neuropathic Pain: emerging treatments. **British Journal of Anaesthesia**, n. 1, v.101, p. 48-58, 2008.

10. MAURICE, T.; SU, T. P.; The Pharmacology of sigma-1 receptors. **Pharmacology and Therapeutics**, v. 124, p. 195-206, 2009.
11. ZAMANILLO, D.; ROMERO, L.; MERLOS, M.; VELA, J. M. Sigma-1 receptor: A new therapeutic target for pain. **European Journal of Pharmacology**, v. 716, p. 78-93, 2013.
12. MARTIN, W. R.; EADES, C. G.; THOMPSON, J.A.; HUPPLER, R. E.; GILBERT, P. E. The effects of morphine- and nalorphine-like drugs in the nondependent and morphine-dependent . **Journal of Pharmacology and Experimental Therapeutics**, v. 197, n. 3, p. 517-532, 1976.
13. HANNER, M.; MOEBIUS, F. F.; FLANDORFER, A.; KNAUS, H.G.; STRIESSNIG, J.; E. KEMPNER, E.; GLOSSMANN, H. Purification, molecular cloning, and expression of the mammalian sigma1-binding site. **Proceedings of the National Academy of Sciences**, v. 93, p. 8072-8077, 1996.
14. HELLEWELL, S. S.; BRUCE, A.; FEINSTEIN, G.; ORRINGER, J.; WILLIAMS, W.; BOWEN, W. D. Rat liver and kidney contain high densities of sigma-1 and sigma-2 receptor: Characterization by ligand binding and photoaffinity labeling. **European Journal of Pharmacology**, v. 1, n. 268, p. 9-18, 1994.
15. BOWEN, W. D. Sigma Receptor: recent advances and new clinical potentials. **Pharmaceutica Acta Helvetiae**, v. 74, p. 211-218, 2000.
16. CRAWFORD, K. W.; BOWEN, W. D. Sigma-2 receptor agonists activate a novel apoptotic pathway and potentiate antineoplastic drugs in breast tumor cell lines. **Cancer Research**, v. 1, p. 313–322, 2002.
17. SU, T. P.; HAYASHI, T. Understanding the Molecular Mechanism of Sigma-1 Receptor: Towards a hypothesis that Sigma-1 Receptors are Intracellular Amplifiers for Signal Transduction. **Current Medicinal Chemistry**, v. 10, p. 2073-2080, 2003

18. HAYASHI, T.; SU, T. P. Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate Ca^{2+} signaling and cell survival. **Cell**, v. 131, p. 596-610, 2007.
19. FONTANILLA, D.; JOHANNESSEN M.; HAJIPOUR A. R.; COZZI N. V.; JACKSON M. B.; RUOHO A. E. The hallucinogen N,N-dimethyltryptamine (DMT) is an endogenous sigma-1 receptor regulator. **Science**, v. 323, p. 934-937, 2009.
20. FISHBACK, J.A.; ROBSON, M. J.; XU, Y.T.; MATSUMOTO R. R. Sigma receptors: potential targets for a new class of antidepressant drugs. **Pharmacology & Therapeutics**, v. 127, p. 271–282, 2010.
21. SU T. P.; HAYASHI T.; MAURICE T.; BUCH S.; RUOHO A.E. The sigma-1 receptor chaperone as an inter-organelle signaling modulator. **Trends in Pharmacological Sciences**, v.31, n.12, p. 557–66, 2010.
22. GLENNON, R. A. Pharmacophore identification for sigma-1 (sigma1) receptor binding: applications of the “deconstruction-reconstruction-elaboration” approach. Mini reviews in medicinal chemistry, v. 5, n. 10, p. 927-940, 2005.
23. DÍAZ, J. L.; CUBERES, R.; BERROCAL, J.; CONTIJOCH, M.; CHRISTMANN, U.; FERNÁNDEZ, A.; PORT, A.; HOLENZ, J.; BUSCHMANN, H.; LAGGNER, C.; SERAFINI, M. T.; BURGUEÑO, J.; ZAMANILLO, D.; MERLOS, M.; VELA, J. M.; ALAMANSA, C. Synthesis and Biological Evaluation of the 1-Arylpyrazole Class of σ_1 Receptor Antagonists: Identifications of 4-{2-[5-Methyl-1-(naphthalene-2-yl)-1H-pyrazol-3-yl]oxy}ethyl}morpholine (S1RA, E52862). **Journal of Medicinal Chemistry**, v. 55, n. 19, p. 8211-8224, 2012.
24. ABADIAS, M.; ESCRICHE, M.; VAQUÉ, A.; SUST, M.; ENCINA, G. Safety, tolerability and pharmacokinetics of single and multiple doses of a novel sigma-1 receptor in three randomized phase I studies. **British Journal of Clinical Pharmacology**, v. 75, n. 1, p. 103-117, 2013.

25. ROH, D. H.; KIM, H. W.; YOON, S. Y.; SEO, H. S.; KWON, Y. B.; KIM, K. W.; HAN, H. J.; BEITZ, A. J.; NA, H. S.; LEE, J. H. Intrathecal Injection of the ζ^1 Receptor Antagonist BD1047 Blocks Both Mechanical Allodynia and Increases in Spinal NR1 Expression during the Induction Phase of Rodent Neuropathic Pain. **Pain Medicine**, v. 109, p. 879-889, 2008.
26. SKUZA, G.; ROGÓZ, Z. Effect of BD 1047, a sigma-1 receptor antagonist, in the animal models predictive of antipsychotic activity. **Pharmacological Reports**, v. 58, n. 5, p. 626-635, 2006.
27. LAN, Y.; CHEN, Y.; CAO, X.; ZHANG, J.; WANG, J.; XU, X.; QIU, Y.; ZHANG, T.; LIU, W.; LIU, B-F; ZHANG, G. Synthesis and Biological Evaluation of Novel Sigma-1 Receptor Antagonists Based on Pyrimidine Scaffold as Agents for Treating. **Journal of Medicinal Chemistry**, v. 57, n. 24, p. 10404-10423, 2014.
28. EROGLU, E., TÜRKMEN, H. A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme CA-II. **Journal of Molecular Graphics and Modelling**, v. 26, p. 701–708, 2007
29. GU, C.G., JIANG, X., JU, X.H., GONG, X.D., WANG, F., BIAN, Y.R., SUN, C. QSARs for congener-specific toxicity of polyhalogenated dibenzo-p-dioxins with DFT and WHIM theory. **Ecotoxicology and Environmental Safety**, v. 72, p. 60–70, 2009.
30. ARULMOZHIRAJA, S., MORITA, M. Structure–activity relationships for the toxicity of polychlorinated dibenzofurans: approach through density functional theory-based descriptors. **Chemical Research in Toxicology**, v. 17, p. 348–356, 2004.
31. PASHA, F.A., SRIVASTAVA, H.K., SINGH, P.P. Comparative QSAR study of phenol derivatives with the help of density functional theory. **Bioorganic & Medicinal Chemistry**, v. 13, p. 6823–6829, 2005.
32. TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploation. **Molecular Informatics**, v. 29, p. 476-488, 2010.

- 33.KO, G. M.; REDDY, A. S.; GARG, R.; KUMAR, S.; HADAEGH, A. R. Computational Modeling Methods for QSAR Studies on HIV-1 Integrase Inhibitors. **Current Computer-Aided Drug Design**, v. 8, p. 255-270, 2012.
- 34.ARAKAWA, M.; HASEGAWA, K.; FUNATSU, K. The recent trend in QSAR Modeling – Variable selection and 3D-QSAR Methods. **Current Computer-Aided Drug Design**, v. 3, p. 254-252, 2007.
- 35.PULIDO, V. A.; GESTAL, M.; MONTEAGUDO, M. C.; RABUÑAL, J. R.; DORADO, J.; MUNTEANU, C. R. Evolutionary Computation and QSAR Research. **Current Computer-Aided Drug Design**, v. 9, p. 206-225, 2013.
- 36.ROY, K.; DAS, R. N. A Review on Principles, Theory and Practices of 2D-QSAR. **Current Drug Metabolism**, v. 15, p. 346-379, 2014.
- 37.TADESCHINI, R.; CONSONNI, V. Handbook of Molecular Descriptors, Wiley–VCH, Weinheim, 2000.
- 38.SABIN, J. R.; TRICKEY, S. B.; APELL, S. P.; ODDERSHEDE, J. Molecular Shape, Capacitance, and Chemical Hardness. **International Journal of Quantum Chemistry**, v. 77, p. 358–366, 2000.
- 39.PEARSON, R. G. The Principle of Maximum Hardness. **Accounts of Chemical Research**, v. 26, p. 250–255, 1993.
- 40.BERKOWITZ, M. Density Functional Approach to Frontier Controlled Reactions. **Journal of the American Chemical Society**, v.109, p. 4823-4825, 1987.
- 41.ZHANG, L.; WAN, J., YANG, G. A DFT-based QSARs study of protoporphyrinogen oxidase inhibitors: phenyl triazolinones. **Bioorganic & Medicinal Chemistry**, v. 12, p. 6183–6191, 2004.
- 42.CHERMETTE, H. Chemical Reactivity Indexes in Density Functional Theory. **Journal of Computational Chemistry**, v. 20, p. 129-154, 1999.

- 43.OKSEL, C.; MA, C. Y.; LIU, J. J.; WILKINS, T.; WANG, X. Z. (Q)SAR modelling of nanomaterial toxicity: A critical review. **Particuology**, v. 21 p. 1–19, 2015.
- 44.HUANG, J.; FAN, X. Why QSAR Fails: An Empirical Evaluation Using Conventional Computational Approach. **Molecular Pharmaceutics**, v. 8, p. 600–608, 2011.
- 45.FERREIRA, M. M. C.; KIRALJ, R. Métodos quimiométricos em relações quantitativas estrutura-atividade (QSAR). In: MONTANARI, C. A. (Ed.). **Química medicinal: métodos e fundamentos em planejamento de fármacos**. São Paulo: Editora da Universidade de São Paulo, 2011. p. 387-453.
- 46.FERREIRA, M. M. C.; MONTANARI, C. A.; GAUDIO, A. C. Seleção de Variáveis em QSAR. **Química Nova**, v. 25, n. 3, p. 439-448, 2002
- 47.MAHALAKSHMI, M.; KALAIVANI, P.; NESALAMAR, E. K. A Review on Genetic Algorithm and its Applications. *International Journal of Computing Algorithm*, v. 2, p. 415-423, 2013.
- 48.KUMAR, M.; HUSIAN, M. Genetic Algorithm: Review and Application. *International Journal of Information Technology and Knowledge Management*, n. 2, v. 2, p. 451-454, 2010.
- 49.EKA'RT, A.; RENNER, G. Genetic algorithms in computer aided design. **Computer-Aided Design**, v.35, p. 709-726, 2003.
- 50.ARROIO, A.; HONÓRIO, K. M.; SILVA, A. B. F. Propriedades Químico-Quânticas Empregadas em Estudos das Relações Estrutura-Atividade. *Química Nova*, v. 33, n. 3, p. 694-699, 2010.
- 51.HAJIMAHDI, Z; RANJBAR, A.; SURATGAR, A. A.; ZARGHI, A. QSAR Study on Anti-HIV-1 Activity of 4-Oxo-1,4-dihydroquinoline and 4-Oxo-4H-pyrido[1,2-a]pyrimidine Derivatives Using SW-MLR, Artificial Neural Network and Filtering Methods. **Iranian Journal of Pharmaceutical Research**, v. 14, p. 69-75, 2015.

52. BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. **Journal of Microbiological Methods**, v. 43, p. 3–31, 2000.
53. ROY, P. P.; ROY, K. On Some Aspects of Variable Selection for Partial Least Squares Regression Models. **QSAR & Combinatorial Science**, v. 27, n. 3, p. 302-313, 2008.
54. ALEXANDER, D. L. J.; TROPSHA, A.; WINKLER, D. A. Beware of R²: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. **Journal of Chemical Information and Modeling**, n. 55, v. 7, p. 1316-1322, 2015.
55. CASTRO, M. M.; MONTANARI, A.; GAUDIO, A. C. Seleção de Variáveis em QSAR. **Química Nova**, v. 25, n. 3, p. 439-448, 2002.
56. GAUDIO, A. C.; ZANDONADE, E. Proposição, Validação e Análise dos Modelos que Correlacionam estrutura Química e Atividade Biológica. **Química Nova**, v. 24, n. 5, p. 658-671, 2001.
57. GARDNER, M. W.; DORLING, S. R. Artificial Neural Networks (The Multilayer Perceptron) – A Review of Applications in the Atmospheric Sciences. **Atmospheric Environment**, v. 32, n. 14/15, p. 2627-2636, 1998.
58. 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**, v. 22, p. 717-727, 2000.
59. BASHEER, M.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design and application. **Journal of Microbiological Methods**, v. 43, p. 3-31, 2000.
60. GAUSSIAN. FRISCH, M. J.; TRUCKS, G. W.; SCHLEGEL, H. B.; SCUSERIA, G.E.; ROBB, M. A.; CHEESEMAN, J. R.; SCALMANI, G.; BARONE, V.; MENNUCCI, B.; PETERSSON, G. A.; NAKATSUJI, H.; CARICATO, M.; LI, X.; HRATCHIAN, H.P.; IZMAYLOV, A. F.; BLOINO,

- J.; ZHENG, G.; SONNENBERG, J. L.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; VREVEN, T.; MONTGOMERY, JR. J. A.; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M.; HEYD, J. J.; BROTHERS, E.; KUDIN, K. N.; STAROVEROV, V. N.; KOBAYASHI, R.; NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A.; BURANT, J. C.; IYENGAR, S.S.; TOMASI, J.; COSSI, M.; REGA, N.; MILLAM, N. J.; KLENE, M.; KNOX, J. E.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERTS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; ZAKRZEWSKI, V. G.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; DAPPRICH, S.; DANIELS, A.D.; FARKAS, O.; FORESMAN, J. B.; ORTIZ, J. V.; CIOŚŁOWSKI, J.; FOX, D. J. Gaussian, Revisão A. 01: Gaussian, 2009. Programa de Computador.
61. LEE, C.; YANG, W.; PARR, R. G.. **Physical Review B**, n. 37, v. 785, 1988.
62. TODESCHINI, R.; CONSONNI, V.; PAVAM, M. **Dragon 2.1**. Milano, 2002.
63. DE OLIVEIRA, D. B.; GAUDIO, A. C.; BuildQSAR: A new computer program for QSAR analysis. Quantitative Structure-Activity Relationships, v. 19, n. 6, p. 599-601, 2001.
64. PIROUETTE. Versão 3.10. [S.l.]: Infometrix, 2002. Programa de computador.
65. MARTINS, J. P. A.; FERREIRA, M. M. C. *QSAR modeling*: um novo pacote computacional *open source* para gerar e validar modelos QSAR. **Química Nova**, v. 36, n. 4, p.554-560, 2013.
66. MATLAB. Versão 7.12. [S.l.]: MathWorks, 2011. Programa de Computador.

67. GOLBRAIKH, A.; TROPSHA, A. Beware of q^2 ! **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 269-276, 2002.
68. TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. **QSAR & Combinatorial Science**, v. 22, n. 1, p. 69-77, 2003.
69. SHARMA, B. K.; PILANIA, P.; SINGH, P.; PRABHAKAR, Y. S. A QSAR study on 2-(4-methylpiperazin-1-yl)quinoxalines as human histamine H₄ receptor ligands. **Journal of Enzyme Inhibition and Medicinal Chemistry**, v. 26, n. 3, p. 412-421, 2011.
70. YUAN, H.; HUANG, J.; CAO, C. Prediction of Skin Sensitization with a Particle Swarm Optimized Support Vector Machine. **International Journal of Molecular Sciences**, v. 10, p. 3273-3254, 2009.
71. FEDOROWICZ, A.; LINGYI, Z.; SINGH, H.; DEMCHUK, E. QSAR Study of Skin sensitization Using Local Lymph Node Assay Data. **International Journal of Molecular Sciences**, v. 5, p. 56-66, 2004.
72. FILHO, E. B. A.; SILVA, J. W. C.; CAVALCANTI, S. C. H. Quantitative-structure-toxicity relationships and molecular highlights about *Aedes aegypti* larvicidal activity of monoterpenes and related compounds. **Medicinal Chemistry Research**, v. 25, n. 10, p. 2171-2178, 2016.
73. KARELSON, M.; LOBNOV, V. S.; KATRITZKY, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. **Chemical Reviews**, v. 96, n. 3, p. 1027-1044, 1996.
74. LESYK, R.; HAVRYLYUK, D.; DEVINYAK, O. 3D-MoRSE descriptors explained. **Journal of Molecular Graphics and Modeling**, v. 54, p. 194-203, 2014.
75. MIHALIC, Z.; NIKOLIC, S.; TRINAISTIC, N. Comparative study of molecular descriptors derived from the distance matrix. **Journal of Chemical Information and Modeling**, v. 32, n. 1, p. 28-37, 1992.

76. SINGH, P. Molecular descriptors in modeling TNF- α converting enzyme (TACE) inhibition activity of 2-(2-aminothiazol-4-yl)pyrrolidine-based tartrate diamides. **Indian Journal of Chemistry**, v. 52, p. 1325-1341, 2013.
77. BASAK, S. C.; NIEMI, G. J.; VEITH, J. D. A Graph-Theoretic Approach to Predicting Molecular Properties. **Mathematical and Computer Modelling**, v. 14, p. 511-516, 1990.

APÊNDICE

Apêndice A. Compostos derivados do esqueleto de pirimidina que atuam como antagonistas do receptor sigma-1¹ e os respectivos valores de pK_i.

Composto	Estrutura Molecular	pK _i
1		7,44
2		7,02
3		6,41
4		6,94
5		6,37
6		6,31
7		7,93
8		7,78
9		7,52

10		6,42
11		6,34
12		6,50
13		6,82
14		7,64
15		7,51
16		7,73
17		7,87
18		7,08
19		7,33

20		7,41
21		7,69
22		7,46
23		6,73
24		7,71
25		7,56
26		6,60
27		6,02
28		8,02
29		8,43

30		8,63
31		7,65
32		7,41
33		6,66
34		7,71
35		7,56
36		6,56
37		6,13
38		7,94
39		8,45

40		8,40
41		7,35
42		7,23
43		5,72
44		8,71
45		8,41
46		8,35
47		8,94
48		9,01
49		8,97

50	<chem>COc1nc(Cl)c(OCCCN2CCCCC2)n1-c1ccc(Cl)c(Cl)c1</chem>	7,05
51	<chem>COc1nc(Cl)c(OCCCN2CCCCC2)n1-c1cccc2ccccc12</chem>	6,99
52	<chem>COc1nc(Cl)c(OCCCN2CCCCC2)n(C)c1</chem>	6,73
53	<chem>COc1nc(Cl)c(OCCCN2CCCCC2)n(C)c1C</chem>	7,23
54	<chem>COc1nc(Cl)c(OCCCN2CCCCC2)n(C)c1C1CC1</chem>	7,40

Função de transferência	Parâmetros estatísticos															
	Neurônios*	EQM	R ²	DMA	Neurônios*	EQM	R ²	DMA	Neurônios*	EQM	R ²	DMA	Neurônios*	EQM	R ²	DMA
<i>purelin-purelin</i>	1	0,011	0,712	0,086	6	0,007	0,800	0,065	11	0,018	0,516	0,106	16	0,020	0,472	0,118
	2	0,012	0,686	0,085	7	0,010	0,741	0,079	12	0,037	0,016	0,145	17	0,024	0,346	0,124
	3	0,010	0,728	0,077	8	0,013	0,639	0,093	13	0,010	0,731	0,084	18	0,032	0,152	0,142
	4	0,011	0,712	0,085	9	0,016	0,582	0,096	14	0,015	0,585	0,098	19	0,049	-0,307	0,175
	5	0,016	0,564	0,100	10	0,010	0,723	0,079	15	0,010	0,727	0,082	20	0,019	0,493	0,107
<i>purelin-logsig</i>	1	0,019	0,489	0,099	6	0,027	0,284	0,129	11	0,027	0,277	0,129	16	0,027	0,283	0,129
	2	0,023	0,377	0,117	7	0,027	0,278	0,129	12	0,025	0,329	0,128	17	0,024	0,348	0,122
	3	0,026	0,301	0,128	8	0,027	0,280	0,130	13	0,023	0,384	0,118	18	0,025	0,326	0,123
	4	0,021	0,440	0,105	9	0,027	0,288	0,127	14	0,026	0,290	0,129	19	0,024	0,349	0,123
	5	0,025	0,341	0,121	10	0,026	0,296	0,129	15	0,024	0,361	0,120	20	0,025	0,323	0,126
<i>logsig-tansig</i>	1	0,011	0,716	0,080	6	0,010	0,744	0,074	11	0,009	0,754	0,077	16	0,009	0,753	0,077
	2	0,009	0,745	0,074	7	0,009	0,747	0,076	12	0,009	0,747	0,076	17	0,009	0,752	0,074
	3	0,009	0,749	0,078	8	0,010	0,732	0,080	13	0,009	0,746	0,077	18	0,010	0,741	0,085
	4	0,009	0,769	0,074	9	0,009	0,750	0,074	14	0,009	0,750	0,073	19	0,010	0,745	0,080
	5	0,010	0,738	0,079	10	0,009	0,766	0,073	15	0,010	0,740	0,080	20	0,009	0,755	0,073
<i>logsig-purelin</i>	1	0,012	0,690	0,084	6	0,010	0,722	0,081	11	0,011	0,709	0,080	16	0,010	0,740	0,082
	2	0,010	0,721	0,081	7	0,010	0,733	0,081	12	0,011	0,703	0,082	17	0,009	0,750	0,076
	3	0,011	0,698	0,082	8	0,010	0,725	0,079	13	0,010	0,743	0,077	18	0,009	0,753	0,074
	4	0,010	0,740	0,078	9	0,011	0,693	0,088	14	0,010	0,744	0,080	19	0,012	0,673	0,084
	5	0,011	0,713	0,079	10	0,011	0,709	0,084	15	0,010	0,735	0,079	20	0,010	0,737	0,074
<i>logsig-logsig</i>	1	0,022	0,396	0,113	6	0,021	0,437	0,108	11	0,020	0,451	0,107	16	0,019	0,490	0,099
	2	0,021	0,446	0,106	7	0,019	0,480	0,101	12	0,020	0,468	0,104	17	0,020	0,469	0,103
	3	0,021	0,440	0,109	8	0,020	0,451	0,103	13	0,020	0,467	0,103	18	0,019	0,482	0,099
	4	0,023	0,369	0,117	9	0,019	0,486	0,101	14	0,021	0,446	0,108	19	0,020	0,469	0,103
	5	0,020	0,473	0,104	10	0,021	0,441	0,108	15	0,019	0,489	0,099	20	0,019	0,488	0,098

*Número de neurônios na camada intermediária