Aline Alves Oliveira

QSAR e dinâmica molecular no estudo de sistemas biomoleculares: predição da
atividade biológica de antagonistas do receptor sigma-1 e simulações de
bicamadas lipídicas

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

Área de concentração: Físico-química

Orientador: Prof. Dr. Albérico Borges Ferreira da
Silva

Exemplar revisado

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

São Carlos

2016

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico para fins de estudo e pesquisa, desde que citada a fonte.

Assinatura: *Aline Alves Oliveira*
Data: *28 de abril de 2016*

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

Oliveira, Aline Alves

QSAR e dinâmica molecular no estudo de sistemas biomoleculares: predição da atividade biológica de antagonistas do receptor sigma-1 e simulações de bicamadas lipídicas. / Aline Alves Oliveira. — São Carlos, 2016.
109 f.

Tese (Doutorado em Físico-Química) — Instituto de Química de São Carlos / Universidade de São Paulo, 2016.
Edição revisada

Orientador: Prof. Dr. Albérico Borges Ferreira da Silva

1. QSAR. 2. Quimiometria. 3. Química quântica. 4. Modelagem molecular. 5. Dinâmica molecular. I. Título.

*Referências Bibliográficas
conferidas pelo SBI/IQSC*

Dedico este trabalho aos meus pais e irmãos,
exemplos de coragem e perseverança.

AGRADECIMENTOS

São muitas as pessoas que contribuíram com suas palavras, conselhos e direcionamentos durante todo o decorrer do curso de Doutorado. Nada mais justo do que registrar meus sinceros agradecimentos, pois todas elas são fundamentais para o meu processo de crescimento e amadurecimento profissional e pessoal.

Primeiramente, gostaria de agradecer a Deus pelo dom da vida e a São Vicente de Paulo e Antônio Frederico Ozanam, meus intercessores e protetores.

Agradeço imensamente ao professor Albérico pela orientação profissional e pessoal, pela paciência, e por compartilhar seus exemplos de vida que foram fundamentais para que eu pudesse prosseguir com fé e segurança.

Aos professores que contribuíram significativamente durante todo o processo de desenvolvimento deste trabalho: Profa. Káthia Honório, Profa. Roseli Romero, Profa. Karen Weber, Prof. Moacyr Comar, Prof. Cláudio Nahum.

Aos colegas, amigos e parceiros do Grupo de Química Quântica e àqueles que já fizeram parte do mesmo, com os quais crescemos e aprendemos juntos: Célio, Estevão, Aldineia, Amanda, Emmanuela, Natieli, Régis, Ronaldo, Rafael, Ana Cristina, Luiz, Guilherme, Laise, Herisson, Genisson, Júlia, Tiago, Gláucio e Ednilson.

Aos funcionários da Química Estrutural e da Secretaria de Pós-Graduação, sempre dispostos a ajudar: Vânia, Joel, Shirlei, Andrea, Sílvia, Gustavo, Gislei.

Ao CNPq pela concessão da bolsa de estudos.

À professora Rommie Amaro da *University of California, San Diego*, por me receber e contribuir com a realização do doutorado sanduíche, aos membros do Amaro Lab que deram suporte à minha estadia e em especial ao Dr. Pekka Postila, que muito contribuiu para o meu crescimento profissional.

À CAPES pela concessão da bolsa de doutorado sanduíche.

A todos os meus amigos, que são joias raras na minha jornada e sem os quais a vida não teria alegria. Em especial aos amigos acadêmicos, com os quais dividimos vários almoços de domingo, cafés da manhã e momentos de distração: Amanda, Giselle, Fernanda, Aldineia, Ana Eliza, Suzy, Karina, Renata, Aline, Lilian, Orlando, Anderson, Dani.

Àqueles que se tornaram amigos de infância em terras americanas e que serão para sempre lembrados com muito carinho: Ana Carolina, Priscila, Hamanda, Bruna, Ítalo, João, André, Rafael. E a todos os amigos *San Diegans*.

E àqueles que são amigos-irmãos na fé e na caridade e que fizeram com que eu tivesse uma grande família em São Carlos: Rafaela, Jocimar, Luciano, Natalia, Cristiano Arthur,

Cristiano Henrique, Aline, Gustavo, Anderson, Gabi, Carlos, Thais, Maurício, Isaías, Fernanda, Tiago, Sebastião, Margareth, e muitos outros companheiros de caminhada vicentina.

Um agradecimento especial ao meu namorado Régis que esteve ao meu lado me dando forças, amor, carinho e tornando meus dias mais felizes, e que também contribuiu muito para a realização deste trabalho com suas sugestões e críticas sempre pertinentes. Agradeço também à minha sogra Maria das Graças e ao meu sogro José Pereira pelas orações e por se preocuparem sempre comigo.

E àqueles que são a razão do meu viver, meus pais Aparecida e Antonio e meus irmãos André Luiz e Alexandre. Sempre me apoiando e fazendo com que eu nunca tivesse medo de seguir adiante, eles são as pessoas que fizeram com que eu chegasse até aqui. Todos os méritos alcançados pertencem a eles também.

*“A mente que se abre a uma nova ideia
jamais voltará ao seu tamanho original”
(Albert Einstein)*

RESUMO

Diferentes abordagens teóricas têm sido utilizadas em estudos de sistemas biomoleculares com o objetivo de contribuir com o tratamento de diversas doenças. Para a dor neuropática, por exemplo, o estudo de compostos que interagem com o receptor sigma-1 (Sig-1R) pode elucidar os principais fatores associados à atividade biológica dos mesmos. Nesse propósito, estudos de Relações Quantitativas Estrutura-Atividade (QSAR) utilizando os métodos de regressão por Mínimos Quadrados Parciais (PLS) e Rede Neural Artificial (ANN) foram aplicados a 64 antagonistas do Sig-1R pertencentes à classe de 1-arilpirazóis. Modelos PLS e ANN foram utilizados com o objetivo de descrever comportamentos lineares e não lineares, respectivamente, entre um conjunto de descritores e a atividade biológica dos compostos selecionados. O modelo PLS foi obtido com 51 compostos no conjunto treinamento e 13 compostos no conjunto teste ($r^2 = 0,768$, $q^2 = 0,684$ e $r^2_{\text{teste}} = 0,785$). Testes de *leave-N-out*, randomização da atividade biológica e detecção de *outliers* confirmaram a robustez e estabilidade dos modelos e mostraram que os mesmos não foram obtidos por correlações ao acaso. Modelos também foram gerados a partir da Rede Neural Artificial Perceptron de Multicamadas (MLP-ANN), sendo que a arquitetura 6-12-1, treinada com as funções de transferência *tansig-tansig*, apresentou a melhor resposta para a predição da atividade biológica dos compostos ($r^2_{\text{treinamento}} = 0,891$, $r^2_{\text{validação}} = 0,852$ e $r^2_{\text{teste}} = 0,793$). Outra abordagem foi utilizada para simular o ambiente de membranas sinápticas utilizando bicamadas lipídicas compostas por POPC, DOPE, POPS e colesterol. Os estudos de dinâmica molecular desenvolvidos mostraram que altas concentrações de colesterol induzem redução da área por lipídeo e difusão lateral e aumento na espessura da membrana e nos valores de parâmetro de ordem causados pelo ordenamento das cadeias acil dos fosfolipídeos. As bicamadas lipídicas obtidas podem ser usadas para simular interações entre lipídeos e pequenas moléculas ou proteínas contribuindo para as pesquisas associadas a doenças como Alzheimer e Parkinson. As abordagens usadas nessa tese são essenciais para o desenvolvimento de novas pesquisas em Química Medicinal Computacional.

Palavras-chave: 1-arilpirazol, Sig-1R, QSAR, PLS, MLP-ANN, POPC, DOPE, POPS, colesterol, membrana sináptica

ABSTRACT

Different theoretical approaches have been used in the studies of biomolecular systems aiming to contribute with the treatment of several diseases. For neuropathic pain, for example, the study of compounds that interact with sigma-1 receptor (Sig-1R) can elucidate the main factors associated to their biological activities. For this purpose, studies of Quantitative Structure-Activity Relationships (QSAR) using Partial Least Squares (PLS) and Artificial Neural Network (ANN) methods were applied to 64 Sig-1R antagonists belong to 1-arylpyrazole class. PLS and ANN models were used in order to describe linear and nonlinear behavior, respectively, between a set of descriptors and the biological activity of the selected compounds. The PLS model was obtained with 51 compounds in the training set and 13 compounds in the test set ($r^2 = 0.768$, $q^2 = 0.684$ and $r^2_{\text{test}} = 0.785$). Leave- N -out tests, biological activity randomization and outliers detection confirmed the robustness and stability of the models and showed that they were not obtained by chance correlations. Models were also generated from Multilayer Perceptron Artificial Neural Network (MLP-ANN) and the 6-12-1 architecture, trained by *tansig-tansig* transfer functions, showed the best result for the biological activity prediction of the compounds ($r^2_{\text{training}} = 0.891$, $r^2_{\text{validation}} = 0.852$ and $r^2_{\text{test}} = 0.793$). Another approach was used to simulate synaptic membranes environment using lipid bilayers composed by POPC, DOPE, POPS and cholesterol. Performed molecular dynamics studies showed that high cholesterol concentration induces decrease of area per lipid and lateral diffusion and increase of membrane thickness and order parameter caused by ordering of phospholipids acil chains. The obtained lipid bilayers can be used to simulate interactions between lipids and small molecules or proteins contributing for researches associated to Alzheimer and Parkinson diseases. The approaches used in this thesis are essential for the development of new researches in Computational Medicinal Chemistry.

Keywords: 1-arylpyrazol, Sig-1R, QSAR, PLS, MLP-ANN, POPC, DOPE, POPS, cholesterol, synaptic membrane

LISTA DE FIGURAS

CAPÍTULO 1

Figura 1 - Representação esquemática da estrutura do Sig-1R.	21
Figura 2 - Estrutura geral da classe de 1-arilpirazóis, antagonistas do Sig-1R. ¹⁰	22
Figura 3 - Representação esquemática de neurônios biológicos (a) e artificiais (b).	27
Figura 4 - Funções de transferência utilizadas no treinamento de ANNs.	28
Figura 5 - Representação de um neurônio artificial.	28
Figura 6 - Coeficientes de regressão para o modelo PLS.	42
Figura 7 - Detecção de <i>outliers</i> no modelo PLS.	42
Figura 8 - Compostos identificados na análise de <i>outliers</i> com valores altos de <i>leverage</i> (compostos 6 e 41) e valores fora dos limites dos resíduos de Student (compostos 40 e 43). Em destaque, os grupos substituintes que se diferenciam dos demais compostos do conjunto treinamento.	43
Figura 9 - Testes de <i>leave-N-out</i> aplicados ao modelo PLS.	44
Figura 10 - Testes de randomização dos valores de pK_i aplicados ao modelo PLS.	44
Figura 11 - Predição da atividade biológica pelo modelo PLS.	45
Figura 12 - Desempenho das MLP-ANNs em relação ao número de neurônios na camada intermediária e às funções de transferência utilizadas.	48
Figura 13 - Diagrama esquemático da MLP-ANN selecionada.	49
Figura 14 - EQM para os conjuntos treinamento, validação e teste em função do número de iterações.	51
Figura 15 - Predição da atividade biológica pelo modelo MLP-ANN.	51

CAPÍTULO 2

Figura 1 - Agregados de lipídeos formados em água: micela, bicamada e lipossoma.	69
Figura 2 - Modelo do mosaico fluido proposto por Singer e Nicolson em 1972. As estruturas sólidas representam as proteínas integrais de membrana que estão distribuídas em uma bicamada lipídica fluida.	70
Figura 3 - Representação estrutural dos lipídeos utilizados nas simulações de bicamadas lipídicas.	82
Figura 4 - APL para o sistema controle em diferentes temperaturas.	86
Figura 5 - Densidade do grupo fosfato das moléculas de POPC em relação ao centro do sistema controle na simulação realizada a 310 K.	87

Figura 6 - DQM das moléculas de POPC no sistema controle simulado nas temperaturas 298, 310 e 323 K.	88
Figura 7 - POs obtidos a partir das simulações por DM do sistema controle em função da variação da temperatura.	89
Figura 8 - Representação do sistema controle simulado a 310 K no tempo de simulação de 300 ns. As esferas vermelhas representam os íons de sódio e as esferas verdes os íons de cloro.	90
Figura 9 - Representação dos sistemas 1, 2 e 3 simulados a 310 K nos tempos de simulação de 0, 100, 200 e 300 ns. As moléculas de POPC, DOPE, POPS e colesterol estão representadas pelas cores azul, amarelo, vermelho e verde, respectivamente.	91
Figura 10 - APLs obtidas a partir das simulações por DM do sistema controle e dos sistemas 1, 2 e 3.	92
Figura 11 - POs obtidos a partir das simulações por DM do sistema controle e dos sistemas 1, 2 e 3.	93
Figura 12 - Posição da molécula de colesterol em relação a moléculas de POPC no sistema 1. A molécula de colesterol está na representação de esferas de van der Waals.	94
Figura 13 - Perfis de densidade obtidos a partir das simulações por DM do sistema controle e dos sistemas 1, 2 e 3.	96

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1 - Representação estrutural e valores de K_i dos compostos da classe de 1-arilpirazóis, antagonistas do Sig-1R (os compostos do conjunto teste estão sublinhados).	31
Tabela 2 - Parâmetros estatísticos do melhor modelo gerado a partir da técnica de seleção de variáveis por algoritmo genético.	39
Tabela 3 - Valores dos descritores selecionados para os antagonistas do Sig-1R (os compostos do conjunto teste estão sublinhados).	39
Tabela 4 - Variância percentual acumulada, SEV, SEC, PRESS, q^2 e r^2 dos modelos PLS.	41
Tabela 5 - Resíduos de predição do conjunto teste feita pelo modelo PLS.	45
Tabela 6 - Parâmetros estatísticos para a avaliação da capacidade preditiva do modelo PLS.	47
Tabela 7 - Melhores modelos de MLP-ANN gerados a partir de diferentes combinações das funções de transferência <i>tansig</i> , <i>logsig</i> e <i>purelin</i>	48
Tabela 8 - Valores de pesos e <i>bias</i> do modelo de MLP-ANN selecionado.	50
Tabela 9 - Resíduos de predição do conjunto teste feita pelo modelo MLP-ANN selecionado.	52

CAPÍTULO 2

Tabela 1 - Modelos de bicamadas lipídicas construídos no servidor CHARMM-GUI ^{59,60} e suas respectivas composições em termos do número de moléculas de lipídeos, água e íons.	82
Tabela 2 - Etapas de equilibração das simulações por DM dos sistemas investigados.	83
Tabela 3 - Valores calculados (<i>Cal.</i>) e experimentais (<i>Exp.</i>) de APL e EB para o sistema controle.	86
Tabela 4 - Valores calculados (<i>Cal.</i>) e experimentais (<i>Exp.</i>) de DL para o sistema controle.	88
Tabela 5 - EB para os sistemas simulados a 310 K.	95
Tabela 6 - DL para os sistemas simulados a 310 K.	97

LISTA DE ABREVIATURAS E SIGLAS

3D-MoRSE	<i>3D Molecule Representation of Structures based on Electron diffraction</i>
ANN	<i>Artificial Neural Network</i>
APL	Área por lipídeo
BiP	<i>Binding immunoglobulin protein</i>
DFT	<i>Density Functional Theory</i>
DL	Difusão lateral
DM	Dinâmica molecular
DMA	Desvio médio absoluto
DMPC	1,2-dimiristoil-sn-glicero-3-fosfocolina
DN	Dor neuropática
DOPE	1,2-dioleoil-sn-glicero-3-fosfoetanolamina
DPPC	1,2-dipalmitoil-sn-glicero-3-fosfocolina
EB	Espessura da bicamada
EQM	Erro quadrático médio
GABA	<i>Gamma-aminobutyric acid</i>
GETAWAY	<i>Geometry, Topology and Atom-Weights Assembly</i>
GGI6	Índice de carga topológica de Gálvez de ordem 6
GPCR	<i>G protein-coupled receptor</i>
H7m	Autocorrelação H de comprimento 7 ponderada pela massa atômica
IASP	<i>International Association for the Study of Pain</i>
IP3	<i>Inositol trisphosphate</i>
MLOGP	Coeficiente de partição octanol-água de Moriguchi
MLP-ANN	<i>Multilayer Perceptron Artificial Neural Network</i>
MM	Mecânica molecular
Mor23m	Sinal 23 dos descritores 3D-MoRSE ponderado pela massa atômica
NMDA	<i>N-methyl-D-aspartate</i>
PLS	<i>Partial Least Squares</i>
PME	<i>Particle-Mesh Ewald</i>
PO	Parâmetro de ordem
POPC	1-palmitoil-2-oleoil-sn-glicero-3-fosfocolina
POPS	1-palmitoil-2-oleoil-sn-glicero-3-fosfoserina

PRESS	<i>Prediction residues error squares sum</i>
QSAR	<i>Quantitative Structure-Activity Relationships</i>
SEC	<i>Standard error of calibration</i>
SEV	<i>Standard error of validation</i>
Sig-1R	<i>Sigma-1 receptor</i>
Sig-2R	<i>Sigma-2 receptor</i>
SRW06	Contagem de caminho de auto-retorno de ordem 6
VL	Variável latente

SUMÁRIO

INTRODUÇÃO GERAL	16
CAPÍTULO 1: Estudos de QSAR de antagonistas do receptor sigma-1	18
1 INTRODUÇÃO	19
1.1 Dor neuropática	19
1.2 Sig-1R e seus ligantes	20
1.3 Métodos computacionais	23
1.3.1 Mínimos Quadrados Parciais – PLS	24
1.3.2 Rede Neural Artificial – ANN	26
2 OBJETIVOS	30
3 METODOLOGIA	31
3.1 Modelos QSAR	36
4 RESULTADOS E DISCUSSÃO	38
4.1 Seleção de descritores	38
4.2 Análises de PLS	40
4.3 Análises de ANN	47
4.4 Proposição de modelos QSAR	52
4.5 Descritores selecionados e a interpretação química dos modelos obtidos	53
5 CONCLUSÕES E PERSPECTIVAS	56
REFERÊNCIAS	57
APÊNDICE	66
CAPÍTULO 2: Proposição de bicamadas lipídicas contendo POPC, DOPE, POPS e colesterol a partir de simulações por dinâmica molecular	68
1 INTRODUÇÃO	69
1.1 Membranas celulares	69
1.1.1 Influência do colesterol nas propriedades físico-químicas das membranas	72

1.1.2 Bicamadas lipídicas e as simulações por DM	73
1.2 Mecânica Molecular	75
1.2.1 Minimização de energia	77
1.2.2 Dinâmica molecular	78
2 OBJETIVOS	81
3 METODOLOGIA.....	82
3.1 Construção das bicamadas lipídicas	82
3.2 Simulações por DM.....	83
4 RESULTADOS E DISCUSSÃO	85
4.1 Construção dos sistemas	85
4.2 Influência da temperatura nas propriedades estruturais do sistema controle	85
4.3 Influência do colesterol nas propriedades estruturais dos sistemas 1, 2 e 3	90
5 CONCLUSÕES E PERSPECTIVAS.....	99
REFERÊNCIAS	100
APÊNDICE	107
CONCLUSÃO GERAL	109

INTRODUÇÃO GERAL

O avanço no desenvolvimento de computadores cada vez mais potentes tem permitido aos pesquisadores em Química Medicinal explorarem metodologias computacionais que muito auxiliam na compreensão de sistemas biomoleculares e no planejamento de moléculas com potencial atividade biológica. Visando redução do tempo de pesquisa em laboratórios experimentais, custos financeiros e melhor compreensão do sistema de interesse, as metodologias computacionais aplicadas na área da Química Medicinal vêm ganhando cada vez mais espaço nas indústrias, universidades e centros de pesquisa.

Quando se pretende investigar, por exemplo, um conjunto de compostos com determinada atividade biológica e que se tem pouca ou nenhuma informação sobre o alvo molecular desses compostos, os estudos das Relações Quantitativas Estrutura-Atividade (do inglês QSAR – *Quantitative Structure-Activity Relationships*) são uma poderosa ferramenta que permite propor modelos matemáticos capazes de prever a atividade biológica desses compostos. Os modelos QSAR possibilitam um planejamento de moléculas a serem testadas experimentalmente, já que os mesmos fornecem indícios de alta ou baixa atividade biológica para as novas moléculas propostas.

Existem ainda outros problemas que podem ser solucionados a partir de simulações por dinâmica molecular (DM) tais como a investigação do comportamento de um determinado sistema ao longo do tempo ou mesmo como as interações entre diferentes moléculas nesse sistema são estabelecidas. As simulações por DM têm sido utilizadas para o estudo de interações entre ligantes e receptores, comportamento de bicamadas lipídicas e até mesmo para simular sistemas maiores como um vírus.

Uma vez que o presente trabalho compreende a utilização dessas duas abordagens para o estudo de sistemas distintos, esta Tese foi dividida em dois capítulos: Capítulo 1 – Estudos de QSAR de antagonistas do receptor sigma-1 e Capítulo 2 – Proposição de bicamadas lipídicas contendo POPC, DOPE, POPS e colesterol a partir de simulações por dinâmica molecular. No Capítulo 1 são propostos modelos de QSAR obtidos a partir das técnicas de Mínimos Quadrados Parciais (do inglês PLS – *Partial Least Squares*) e Rede Neural Artificial (do inglês ANN – *Artificial Neural Network*) para um conjunto de 64 moléculas que apresentam atividade biológica para o receptor sigma-1 (do inglês Sig-1R – *Sigma-1 receptor*). Os estudos apresentados foram desenvolvidos no Grupo de Química Quântica do Instituto de Química de São Carlos da Universidade de São Paulo sob a orientação do Prof.

Dr. Albérico Borges Ferreira da Silva. Já no Capítulo 2 são propostos modelos de bicamadas lipídicas a partir de simulações por DM e que apresentam diferentes concentrações de colesterol, se assemelhando à composição lipídica de membranas sinápticas. Os estudos apresentados no Capítulo 2 foram desenvolvidos durante o estágio sanduíche no Amaro Lab na University of California, San Diego, EUA, sob a supervisão da Prof. Dra. Rommie Elizabeth Amaro.

CAPÍTULO 1: Estudos de QSAR de antagonistas do receptor sigma-1

1 INTRODUÇÃO

1.1 Dor neuropática

Segundo a Associação Internacional para o Estudo da Dor (do inglês IASP - *International Association for the Study of Pain*), a dor neuropática (DN) compreende lesões ou doenças do sistema somatossensorial, cujas diferentes causas permitem classificá-la como central ou periférica.¹ A DN periférica surge devido a doenças ou lesões nos nervos periféricos, podendo-se citar como exemplo a neuropatia periférica diabética, neuralgia pós-herpética, neuropatia induzida por terapia antineoplásica ou HIV, dentre outras. Já a DN central ocorre devido a doenças ou lesões do sistema nervoso central, como esclerose múltipla, lesão da medula espinhal, doença de Parkinson ou acidente vascular cerebral.^{2,3}

A DN é diagnosticada a partir de anormalidades sensoriais que incluem a perda de sensação de dor ou mesmo um aumento da sensibilidade. Os pacientes com DN descrevem os sintomas de formas muito diferentes, sendo que as principais descrições dadas por pacientes englobam os termos choque elétrico, ardor e formigamento.⁴

Estudos apontam que 6 a 8% da população de países europeus sofre com dor crônica de origem neuropática.^{5,6} Outras pesquisas, realizadas no Reino Unido, estimam a prevalência de DN em 1% da população.⁷ Entretanto, essas estimativas podem estar equivocadas uma vez que os estudos podem considerar mais vezes o mesmo indivíduo que apresente DN causada por diferentes doenças, ou mesmo restringir as causas de DN.² Portanto, não se tem uma estimativa exata do número de pessoas que sofrem com a DN, mas sabe-se que esse tipo de dor reduz significativamente a qualidade de vida dessas pessoas tornando-se, portanto, cada vez mais necessário o desenvolvimento de novos fármacos para o tratamento.

O tratamento da DN movimentou cerca de 2,4 bilhões de dólares no ano de 2010 nos sete maiores mercados mundiais: Estados Unidos, Japão, França, Alemanha, Itália, Espanha e Reino Unido. Estima-se que em 2020 esse valor chegue a 3,6 bilhões.⁸ Entretanto, o tratamento da DN é muito complexo. Uma vez que não existe cura, pesquisadores buscam desenvolver tratamentos capazes de impedir o seu progresso melhorando a qualidade de vida do paciente. Além de se utilizar medicamentos, muitas vezes são necessárias intervenções neurocirúrgicas, tratamento psiquiátrico e psicológico e terapias, dentre outros variados tratamentos que podem ajudar no convívio com a DN.³

Uma vez que a DN pode ser oriunda de diferentes causas, os medicamentos que são usados no tratamento são muito variados e incluem analgésicos, anti-inflamatórios, anticonvulsivantes, antidepressivos tricíclicos, antagonistas do receptor N-metil-D-aspartato (do inglês NMDA – *N-methyl-D-aspartate*), analgésicos opióides, canabinóides, dentre outros.^{3,9}

Alguns medicamentos já estão aprovados para o tratamento da DN, enquanto muitos estão em fase de testes. Porém, poucos pacientes apresentam alívio de mais que 50% da dor. O tratamento padrão para a DN utiliza o anticonvulsivante pregabalina (Lyrica[®]), capaz de modular os canais de cálcio, além de apresentar atividade agonista ao ácido γ -aminobutírico (do inglês GABA – *Gamma-aminobutyric acid*). Também se tem utilizado outros medicamentos como o antidepressivo duloxetina (Cymbalta[®]), o anticonvulsivante gabapentina (Gralise[®]) e o tapentadol (Nucynta ER[®]) que atua como agonista do receptor μ -opioide e também inibe a recaptação de noradrenalina e também os medicamentos Lidoderm[®] e Qutenza[®].⁸

Diante do exposto, torna-se imprescindível a busca constante por alvos mais apropriados para que novos fármacos sejam desenvolvidos para o tratamento da DN. Pesquisas recentes têm apontado o Sig-1R como um desses promissores alvos. Compostos de diferentes classes capazes de se ligar ao Sig-1R, tais como: derivados de 1-arilpirazol ou mesmo do adamantano, têm sido desenvolvidos e apresentam grande potencial para o tratamento da DN.¹⁰⁻¹²

1.2 Sig-1R e seus ligantes

Descoberto em 1976 por Martin e colaboradores, o Sig-1R foi inicialmente classificado como um novo receptor opioide. Essa classificação foi dada em função de um estudo da farmacologia de compostos agonistas, a morfina e análogos, que foi capaz de identificar três receptores (μ , κ e σ) que interagiam com essas moléculas.¹³ Contudo, foi observado que os efeitos dos compostos capazes de interagir com o Sig-1R não eram antagonizados por naloxona e naltrexona, antagonistas de receptores opioides.^{14,15}

Atualmente, receptores do tipo sigma são considerados uma classe particular de receptores e englobam dois subtipos: Sig-1R e receptor sigma-2 (do inglês Sig-2R – *Sigma-2 receptor*). O Sig-1R é uma proteína encontrada predominantemente no retículo endoplasmático, sendo constituído por 223 aminoácidos e com massa molecular de 25

kDa.^{16,17} Foi clonado pela primeira vez em 1996, sendo o receptor do tipo sigma mais amplamente estudado.¹⁸ Até o momento, março de 2016, não se tem dados cristalográficos deste receptor, pois se trata de uma proteína de membrana cuja cristalização é de difícil obtenção. Esse fato dificulta o desenvolvimento de novos e potentes ligantes do Sig-1R. Entretanto, já foram construídos modelos por homologia para obtenção de sua estrutura tridimensional, auxiliando no planejamento de novos fármacos.¹⁹⁻²¹

Acredita-se que a estrutura do Sig-1R possua dois domínios transmembranares (resíduos 10-30 e 80-100) ligados por um *loop* extracelular, sendo que as extremidades N-terminal e C-terminal estão localizadas na região intracelular (Figura 1).²² Estudos mostram ainda que a extremidade C-terminal (resíduos 116-223) exerce a função de chaperona molecular, ou seja, é capaz de auxiliar no processo de enovelamento de proteínas.^{23,24}

Figura 1 - Representação esquemática da estrutura do Sig-1R.

Fonte: Adaptação a partir de HAYASHI, T. *et al.*, 2007, p. 596-610.

A atividade de chaperona do Sig-1R envolve a interação física com a proteína de ligação de imunoglobulinas (do inglês BiP – *binding immunoglobulin protein*), outra chaperona do retículo endoplasmático. Quando ocorre a formação de um complexo entre o Sig-1R e a BiP, a atividade de chaperona é minimizada e quando o complexo é desfeito, ocorre a maximização da atividade de chaperona. Compostos sintéticos que atuam como agonistas do Sig-1R são capazes de promover a dissociação entre Sig-1R e BiP, aumentando assim a atividade de chaperona do Sig-1R, enquanto que compostos antagonistas reforçam a

associação entre eles, diminuindo a atividade de chaperona do Sig-1R. Íons Ca^{2+} também são capazes de interferir na associação das chaperonas reforçando a interação em concentrações fisiológicas. A associação entre Sig-1R e BiP, regulada por íons Ca^{2+} , é controlada por receptores acoplados a proteína G (do inglês GPCR – *G protein-coupled receptor*). Receptores de inositol trifosfato (do inglês IP3 – *Inositol trisphosphate*), ativados pelo estímulo de GPCRs, promovem uma diminuição de íons Ca^{2+} , contribuindo para a dissociação entre Sig-1R e BiP.²³

Os receptores do tipo sigma estão amplamente distribuídos no sistema nervoso central e em tecidos periféricos atuando em diversas funções celulares como regulação e modulação de canais iônicos, além de estarem associados à sobrevivência ou morte de células.^{25–27} Portanto, eles têm sido estudados como possíveis alvos para o tratamento de doenças como DN, depressão e câncer.^{26,28}

Diferentes classes de compostos têm sido propostas como possíveis agonistas ou antagonistas do Sig-1R.^{11,29} Recentemente, Díaz e colaboradores sintetizaram uma classe de 1-arilpirazóis, antagonistas do Sig-1R (Figura 2).¹⁰ Neste estudo a seletividade dos compostos em relação ao Sig-1R e ao Sig-2R foi avaliada, levando à seleção do composto S1RA, que atualmente se encontra em fase de estudos clínicos.^{30,31} Os resultados mostram que a atividade farmacológica desses compostos está relacionada ao tipo de substituinte e que o espaçador etilenoxi e pequenas aminas cíclicas foram essenciais para uma maior seletividade.¹⁰

Figura 2 - Estrutura geral da classe de 1-arilpirazóis, antagonistas do Sig-1R.¹⁰

Uma vez que é necessário desenvolver novos compostos para o tratamento da DN, um estudo mais amplo de compostos já sintetizados é capaz de permitir um maior entendimento dos principais fatores associados à atividade biológica dos mesmos. Nesse propósito, estudos teóricos utilizando métodos computacionais aplicados aos compostos, como aqueles da classe de 1-arilpirazóis, pode contribuir para o planejamento de novos antagonistas do Sig-1R e, conseqüentemente, para o desenvolvimento de novos fármacos para o tratamento da DN.

1.3 Métodos computacionais

O expressivo desenvolvimento dos computadores nas últimas décadas tem proporcionado avanços significativos em diversos campos científicos, sendo a Química Medicinal muito beneficiada por este processo. O custo financeiro e o tempo associados ao desenvolvimento de um fármaco são muito altos, porém todo o processo pode ser acelerado quando ferramentas computacionais são utilizadas.^{32,33}

Dentre as diversas técnicas computacionais que podem ser empregadas no planejamento de fármacos, estudos de QSAR têm sido amplamente utilizados permitindo desenvolver modelos matemáticos capazes de descrever a atividade biológica de um grupo de compostos levando em consideração suas características estruturais.³⁴

Uma vez que é possível obter diversas variáveis capazes de descrever características estruturais e eletrônicas de um conjunto de compostos, torna-se necessário utilizar diferentes técnicas de seleção dessas variáveis para posteriores estudos de QSAR. Dessa forma, o conjunto inicial de variáveis será reduzido, permitindo eliminar aquelas que forem menos relevantes, antes mesmo da geração dos modelos de QSAR. Dentre as diferentes técnicas de seleção dessas variáveis, pode-se utilizar a análise de valores de correlação com variáveis medidas experimentalmente, tal como a atividade biológica. Recomenda-se que aquelas variáveis que possuam valores de correlação menores que 0,3 devem ser excluídas. Também é possível investigar a correlação entre as variáveis, sendo que aquelas que foram altamente correlacionadas entre si devem ser removidas.³⁵

Outras metodologias mais elaboradas permitem utilizar técnicas como a busca sistemática ou o algoritmo genético para realizar uma busca pelas variáveis mais importantes. Na busca sistemática são feitas todas as combinações possíveis entre as variáveis, sendo que as melhores regressões lineares múltiplas encontradas para um conjunto de k variáveis serão selecionadas.³⁶ Já no algoritmo genético, a seleção das variáveis é feita por três processos evolutivos: seleção natural, cruzamento e mutação.³⁷

Uma vez selecionadas as variáveis mais relevantes, os modelos QSAR podem ser obtidos e validados. Nesse propósito, métodos de regressão tais como PLS e ANN têm sido aplicados a diferentes conjuntos de moléculas, sempre conduzindo a modelos com alta capacidade preditiva.³⁸⁻⁴¹ Comparações entre os modelos gerados por estes métodos têm permitido identificar se um determinado conjunto de variáveis apresenta comportamento

linear ou não linear em relação a uma variável que se deseja modelar.⁴²⁻⁴⁴ Essas informações são fundamentais para uma melhor descrição dos sistemas em estudo.

1.3.1 Mínimos Quadrados Parciais – PLS

A regressão por PLS descreve relações lineares entre um conjunto de descritores e uma determinada medida experimental, tal como valores de atividade biológica. Neste método, os descritores (matriz X) são denominados de variáveis independentes enquanto que a medida experimental (matriz Y) constitui a variável dependente.

Na PLS são obtidas as variáveis latentes (VLs) que representam combinações lineares das variáveis originais das matrizes X ou Y . As VLs são obtidas e selecionadas de forma que descrevam a máxima quantidade de variância das amostras.

As matrizes X e Y são decompostas simultaneamente em função da soma das a VLs e das matrizes de erros que correspondem a parte não modelada, tal como nas equações 1 e 2:

$$X = TP' + E = \sum t_a p'_a + E \quad (1)$$

$$Y = UQ' + F = \sum u_a q'_a + F \quad (2)$$

em que T e U representam as matrizes de escores, P' e Q' representam as matrizes de pesos, E e F são 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' , respectivamente.

Para se estabelecer uma relação entre as matrizes X e Y , os escores podem ser correlacionados utilizando a matriz B que contém os coeficientes de regressão (eq. 3):

$$U = BT \quad (3)$$

Portanto, a matriz Y pode ser reescrita da seguinte forma (eq. 4):

$$Y = TBQ' + F \quad (4)$$

Testes de validação cruzada com a exclusão de N compostos (*leave-N-out*) podem ser aplicados aos modelos PLS para determinar valores dos erros de predição. Na validação

cruzada são removidas algumas amostras do conjunto de calibração que serão posteriormente utilizadas para realizar a predição e o cálculo dos resíduos. Esse procedimento é repetido até que todas as amostras tenham sido utilizadas para realizar a predição.⁴⁵⁻⁴⁷

A análise dos parâmetros estatísticos gerados pelos modelos PLS permite definir um número ótimo de VLs. Recomenda-se que este procedimento seja feito buscando-se os menores valores dos erros de predição. Dentre os parâmetros estatísticos devem-se avaliar os valores da soma dos quadrados dos erros de predição (do inglês PRESS – *Prediction Residues Error Squares Sum*), erro padrão de validação (do inglês SEV – *Standard Error of Validation*), erro padrão de calibração (do inglês SEC – *Standard Error of Calibration*), q^2 (coeficiente de correlação da validação cruzada) e r^2 (coeficiente de determinação).^{35,48} As equações destes parâmetros estão descritas a seguir (eqs. 5 a 9):

- PRESS

$$PRESS = \sum (y_{iexp} - \hat{y}_{ival})^2 \quad (5)$$

em que y_{iexp} representa o valor experimental de Y da i -ésima amostra e $\hat{y}_{ival}$ representa o valor predito de Y pelo modelo construído sem a amostra i .

- SEV

$$SEV = \left[\frac{PRESS}{I} \right]^{\frac{1}{2}} \quad (6)$$

sendo que I representa o número de compostos no conjunto treinamento utilizados na validação cruzada.

- SEC

$$SEC = \left[\frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{I - a - 1} \right]^{\frac{1}{2}} \quad (7)$$

sendo que $\hat{y}_{ical}$ representa o valor calculado de Y da i -ésima amostra utilizando o modelo construído com todas as amostras.

- q^2

$$q^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - y_{mean})^2} \quad (8)$$

em que y_{mean} representa o valor experimental médio de Y .

- r^2

$$r^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - y_{mean})^2} \quad (9)$$

1.3.2 Rede Neural Artificial – ANN

O método ANN também pode ser aplicado a um conjunto de descritores com o propósito de propor modelos para a predição da atividade biológica. Com esse método é possível investigar se os descritores apresentam algum tipo de relação não linear com a medida experimental investigada, além de permitir obter modelos altamente preditivos.⁴⁹

As ANNs são inspiradas no funcionamento de neurônios biológicos e projetadas para processar informações e adquirir conhecimento. Em um neurônio artificial há uma região de processamento da informação que possui duas extremidades: entradas e saída. As entradas são como os dendritos em um neurônio biológico, enquanto que a saída representa o axônio. Portanto, a região de processamento corresponde ao corpo celular (Figura 3).

Figura 3 - Representação esquemática de neurônios biológicos (a) e artificiais (b).

Em um neurônio biológico o sinal entra pelos dendritos, passa pelo corpo celular e, através do axônio, é transmitido para outros neurônios. Da mesma forma acontece em um neurônio artificial, onde as entradas são ponderadas por um fator, ou peso sináptico, antes de chegarem à região de processamento, assim como acontece nas sinapses dos neurônios biológicos. Os sinais de entrada multiplicados pelos pesos sinápticos serão somados e uma determinada função de transferência será aplicada fornecendo o sinal da saída. As equações 10 e 11 apresentam a descrição matemática de um neurônio artificial n :

$$u_n = \sum_{i=1}^m w_{ni} \cdot x_i \quad (10)$$

em que u_n corresponde ao sinal que entra na região de processamento do neurônio n , m é o número de sinais de entrada no neurônio n , w_{ni} representa os pesos sinápticos e x_i , os sinais de entrada.

$$y_n = f(u_n) \quad (11)$$

sendo que f representa a função de transferência utilizada e y_n , o sinal de saída do neurônio n .

As funções de transferência são capazes de conferir a não linearidade a uma ANN, atuando sobre as entradas que chegam no neurônio. Existem diferentes funções de transferência, tais como as funções *logsig* (função de transferência logarítmica sigmoidal) e *tansig* (função de transferência tangente hiperbólica sigmoidal). Além dessas funções, existe também uma função linear representada pela *purelin* (função de transferência linear). A Figura 4 mostra uma representação esquemática das funções de transferência mencionadas.

Figura 4 - Funções de transferência utilizadas no treinamento de ANNs.

Os neurônios artificiais são então interligados para formar uma ANN e sua arquitetura define o número de neurônios que estão presentes na camada de entrada, intermediária e de saída. Portanto, a saída final de uma ANN será obtida calculando-se de forma sequencial as entradas e saídas por toda a rede. Pode-se ainda acrescentar a uma ANN um termo independente dos dados de entrada, *bias*, que é utilizado para otimizar o ajuste da mesma. Um esquema simplificado de um neurônio artificial pode ser visualizado na Figura 5.

Figura 5 - Representação de um neurônio artificial.

A disposição dos neurônios em uma ANN pode compreender três tipos principais de redes: alimentadas adiante, redes recorrentes e redes competitivas. Nas primeiras, os neurônios estão dispostos em camadas e as informações passam pela camada de entrada até atingirem a camada de saída. Nas redes recorrentes a saída de um neurônio é utilizada como entrada no próprio neurônio ou em neurônios de camadas anteriores. Já nas redes competitivas, há uma competição entre os neurônios da camada de saída sendo que o vencedor será ativado a cada iteração.

Existem ainda diferentes modelos de ANNs que devem ser escolhidos de acordo com o objeto de estudo. Dentre esses modelos, as ANNs do tipo Perceptron de multicamadas (do inglês: MLP-ANN – *Multilayer Perceptron Artificial Neural Network*) são amplamente utilizadas. A MLP-ANN é um tipo de rede alimentada adiante que possibilita a inserção de camadas intermediárias entre as camadas de entrada e de saída.

O treinamento de uma MLP-ANN permite que a mesma possa modelar um conjunto de dados ou mesmo reconhecer algum padrão associado a esses dados. Para isso, a MLP-ANN obtém valores de erros de predição que serão empregados na otimização dos pesos sinápticos. Tal otimização pode ser feita utilizando algoritmos como a retropropagação de erros. Dessa forma, a correção dos valores dos pesos sinápticos se inicia na última camada e continua até atingir a primeira camada.^{50,51}

Coefficientes estatísticos como coeficiente de determinação (r^2 ; eq. 9), erro quadrático médio (EQM) e desvio médio absoluto (DMA) podem ser empregados no processo de avaliação de uma ANN⁵⁰. As equações 13 e 14 descrevem como os coeficientes EQM e DMA são calculados:

$$EQM = \frac{1}{N} \sum_{i=1}^N (y_{ical} - y_{iexp})^2 \quad (13)$$

$$DMA = \frac{1}{N} \sum_{i=1}^N |y_{ical} - y_{iexp}| \quad (14)$$

2 OBJETIVOS

O presente estudo consiste na proposição de modelos QSAR capazes de prever a atividade biológica dos compostos da classe 1-arilpirazóis, antagonistas do Sig-1R. Os objetivos específicos compreendem:

- Obtenção e seleção de descritores moleculares e eletrônicos para os compostos investigados;
- Proposição e validação de modelos QSAR utilizando o método PLS;
- Proposição e validação de modelos QSAR utilizando o método ANN;
- Escolha dos melhores modelos para predição da atividade biológica de novos compostos pertencentes à classe de 1-arilpirazóis.

3 METODOLOGIA

Estudos de QSAR foram aplicados aos compostos da classe de 1-arilpirazóis sintetizados por Díaz e colaboradores.¹⁰ Para o presente estudo foram selecionados os compostos com valores de constante de inibição (K_i) para o Sig-1R determinados experimentalmente utilizando membranas transfectadas de rim embrionário humano (HEK-293) e a molécula de [^3H](+)-pentazocina como radioligante.¹⁰ Os compostos foram divididos em conjunto treinamento e conjunto teste, de forma que a faixa de variação dos valores de K_i fosse adequadamente representada no conjunto teste, considerando ainda a diversidade química dos compostos. Para cada intervalo de variação de 0,5 unidades de pK_i ($-\log K_i$) foram retirados, em média, de 20 a 30% dos compostos para constituírem o conjunto teste. Na Tabela 1 são apresentadas as estruturas dos compostos selecionados, bem como os respectivos valores de K_i .

Tabela 1 - Representação estrutural e valores de K_i dos compostos da classe de 1-arilpirazóis, antagonistas do Sig-1R (os compostos do conjunto teste estão sublinhados).

Comp.	Estrutura Geral	R ₁	R ₂	R ₃	X	Y	Z	K_i (nm) ¹⁰	pK_i
<u>1</u>		Me		-	-	-	-	<u>10,8 ± 1,8</u>	<u>7,967</u>
<u>2</u>		Me		-	-	-	-	<u>5,2 ± 2,3</u>	<u>8,284</u>
3		Me		-	-	-	-	5,5 ± 1,2	8,260
4		Me		-	-	-	-	9,4 ± 0,3	8,027
5		Me		-	-	-	-	8,4 ± 1,0	8,076
6		Me		-	-	-	-	16,0 ± 1,2	7,796
7		Me		-	-	-	-	11,1 ± 0,5	7,955
8		Me		-	-	-	-	44,0 ± 0,9	7,357

Comp.	Estrutura Geral	R ₁	R ₂	R ₃	X	Y	Z	K _i (nm) ¹⁰	pK _i
9		Me		-	-	-	-	14,8 ± 3,7	7,830
10		Me		-	-	-	-	9,9 ± 0,3	8,004
11		Me		-	-	-	-	8,5 ± 0,2	8,071
<u>12</u>		Me		-	-	-	-	294 ± 29	6,532
13		H		-	-	-	-	8,2 ± 0,7	8,086
14		H		-	-	-	-	15,7 ± 2,7	7,804
15		H		-	-	-	-	16,7 ± 1,2	7,777
16		H		-	-	-	-	8,9 ± 0,9	8,051
<u>17</u>		H		-	-	-	-	8,6 ± 1,2	8,066
18		H		-	-	-	-	7,2 ± 1,1	8,143
19		Pr ⁱ		-	-	-	-	172 ± 32	6,764
20		Ph		-	-	-	-	224 ± 1	6,650
21		Me		-	(CH ₂) ₃	-	-	18,9 ± 1,6	7,724
<u>22</u>		Me		-	(CH ₂) ₃	-	-	10,4 ± 1,8	7,983
23		Me		-	(CH ₂) ₃	-	-	13,3 ± 0,1	7,876
24		H		-	(CH ₂) ₃	-	-	16,1 ± 3,0	7,793

Comp.	Estrutura Geral	R ₁	R ₂	R ₃	X	Y	Z	K _i (nm) ¹⁰	pK _i
25		H		-	(CH ₂) ₃	-	-	9,6 ± 1,4	8,018
26		H		-	(CH ₂) ₃	-	-	16,5 ± 6,8	7,783
27		Me		-	(CH ₂) ₄	-	-	13,3 ± 0,2	7,876
28		Me		-	(CH ₂) ₄	-	-	6,9 ± 0,1	8,161
29		Me		-	(CH ₂) ₄	-	-	7,1 ± 0,3	8,149
30		Me		-	(CH ₂) ₄	-	-	26,0 ± 3,8	7,585
31		Me		-	(CH ₂) ₄	-	-	9,9 ± 0,5	8,004
32		Me		-	(CH ₂) ₄	-	-	11,7 ± 0,1	7,932
33		Me		-	(CH ₂) ₄	-	-	854 ± 144	6,069
34		H		-	(CH ₂) ₄	-	-	17,6 ± 0,1	7,754
35		H		-	(CH ₂) ₄	-	-	10,2 ± 0,5	7,991
36		H		-	(CH ₂) ₄	-	-	14,8 ± 0,5	7,830
37		H		-	(CH ₂) ₄	-	-	17,2 ± 6,1	7,764
38		H		-	(CH ₂) ₄	-	-	24,0 ± 0,1	7,620
39		H		-	(CH ₂) ₄	-	-	8,2 ± 0,9	8,086

Comp.	Estrutura Geral	R ₁	R ₂	R ₃	X	Y	Z	K _i (nm) ¹⁰	pK _i
40		H		-	(CH ₂) ₂ O(CH ₂) ₂	-	-	36,8 ± 0,1	7,434
41		H		-	(CH ₂) ₂ OSO ₂ (CH ₂) ₂	-	-	879 ± 149	6,056
42		H		-	CH ₂ CHOHCH ₂	-	-	128 ± 9	6,893
43		H		-	CH ₂ CHOHCH ₂	-	-	233 ± 52	6,633
44			Me	Me	-	-	-	9,4 ± 1,8	8,027
45			Me	COMe	-	-	-	741 ± 134	6,130
46			Me	H	-	-	-	180 ± 28	6,745
47			H	H	-	-	-	17,8 ± 1,5	7,750
48			Me	H	-	-	-	44,0 ± 11,0	7,357
49			Me	H	-	-	-	17,0 ± 7,0	7,770
50			H	H	-	-	-	33,3 ± 4,0	7,478
51			Me	Me	-	-	-	139 ± 9	6,857
52			H	H	-	-	-	23,1 ± 3,9	7,636
53			-	-	(CH ₂) ₂	-	-	10,0 ± 0,7	8,000
54			-	-	(CH ₂) ₂	-	-	9,9 ± 0,4	8,004
55			-	-	(CH ₂) ₂	-	-	47,0 ± 1,4	7,328

Comp.	Estrutura Geral	R ₁	R ₂	R ₃	X	Y	Z	K _i (nm) ¹⁰	pK _i
56			-	-	(CH ₂) ₂	-	-	33,8 ± 4,8	7,471
57			-	-	(CH ₂) ₂	-	-	83,4 ± 3,7	7,079
58			-	-	(CH ₂) ₄	-	-	99,1 ± 19,5	7,004
59			-	-	(CH ₂) ₄	-	-	65,0 ± 21,2	7,187
60			-	-	(CH ₂) ₄	-	-	296 ± 15	6,529
61		Me	-	-	S	H ₂	O	14,2 ± 4,4	7,848
62		H	-	-	NH	O	CH ₂	80,8 ± 30,2	7,093
63		H	-	-	NH	H ₂	O	39,4 ± 6,4	7,405
64		H	-	-	NH	H ₂	CH ₂	9,5 ± 4,7	8,022

FONTE: DÍAZ, J. L.; CUBERES, R.; BERROCAL, J.; *et al.*, 2012, p. 8211-8224.

As estruturas tridimensionais dos compostos foram construídas e otimizadas utilizando o pacote Gaussian 09.⁵² Cálculos de otimização de geometria e de frequências vibracionais foram realizados utilizando a Teoria do Funcional de Densidade (do inglês DFT – *Density Functional Theory*) com funcional B3LYP e função de base 6-31++G(d,p). Os descritores eletrônicos foram extraídos dos resultados desses cálculos e as estruturas otimizadas dos compostos foram utilizadas para gerar os descritores moleculares através do programa Dragon 5.⁵³

A fim de realizar uma seleção prévia dos descritores moleculares, os mesmos foram submetidos a análises do coeficiente de correlação (r) entre os valores de cada descritor e os valores de pK_i. Aqueles descritores que apresentaram valores de coeficiente de correlação menores que 0,3 (em módulo) foram excluídos.

A técnica de algoritmo genético para seleção de variáveis implementada no programa BuildQSAR 2.1.0.0⁵⁴ foi utilizada para realizar uma nova seleção dos descritores. Foram gerados diferentes modelos de regressão linear múltipla a partir da variação do número de descritores por modelo, do número de gerações e do critério de corte para a correlação entre os descritores. Após as análises por algoritmo genético foram selecionados os descritores a

serem utilizados nas análises de PLS e ANN, baseando-se em valores de r e de q^2 dos modelos gerados.

3.1 Modelos QSAR

O programa Pirouette⁵⁵ foi utilizado para realizar as análises de PLS. As mesmas foram aplicadas ao conjunto treinamento (51 compostos), que foi validado externamente com o conjunto teste (13 compostos). Para realizar tais análises todos os dados foram inicialmente autoescalados, isto é, os dados foram centrados na média e divididos pelo respectivo desvio padrão.

A seleção dos melhores modelos de PLS baseou-se nos parâmetros estatísticos SEV, SEC, PRESS, q^2 e r^2 . Também foram feitas análises de detecção de *outliers* a fim de identificar amostras com comportamento atípico. Validações adicionais dos modelos foram feitas por meio de testes *leave-N-out* e randomização da atividade biológica utilizando-se o programa QSAR Modeling.⁵⁶ Os testes de *leave-N-out* foram realizados a fim de comprovar a robustez dos modelos gerados, baseando-se nos desvios dos valores de q^2 obtidos. Neste teste foram excluídos N compostos do conjunto treinamento sendo que para cada N , as análises foram repetidas 30 vezes. O valor de N foi definido de acordo com o número de compostos presentes no conjunto treinamento, correspondendo de 20 a 30% dos compostos estudados. A fim de verificar se os modelos foram obtidos a partir de correlações ao acaso, foram realizados 50 testes de randomização dos valores de atividade biológica, sendo realizados testes de *leave-one-out* para cada randomização. Nestes testes analisou-se os valores de r^2 e q^2 após a randomização dos valores de pK_i .

Os modelos de ANN foram treinados utilizando o programa Matlab R2011a.⁵⁷ Utilizou-se MLP-ANNs, constituídas de uma camada de entrada, uma camada intermediária e uma camada de saída. Os neurônios da camada de entrada representam os descritores previamente selecionados, enquanto que os neurônios da camada de saída representam os valores preditos para o pK_i . Todos os compostos selecionados neste estudo foram utilizados para treinar a MLP-ANN usando o algoritmo de retropropagação. Os dados foram escalados em uma faixa de 0,1 a 0,9 baseando-se na normalização de valores máximos e mínimos conforme a equação 15.

$$i' = 0,8 \frac{i - i_{min}}{i_{max} - i_{min}} + 0,1 \quad (15)$$

em que i e i' representam os valores não normalizados e normalizados para as entradas da ANN, respectivamente, e i_{min} e i_{max} são os valores mínimos e máximos para essas entradas.

Para realizar o treinamento das MLP-ANNs, manteve-se o mesmo conjunto teste utilizado nos estudos de PLS. Foram selecionados ainda seis compostos (24, 28, 33, 47, 54, 63) pertencentes ao conjunto treinamento utilizado nos modelos de PLS para comporem o conjunto validação. Para cada intervalo de variação de 1 unidade de pK_i foram retirados, em média, 10% dos compostos. O conjunto validação é necessário durante o processo de treinamento de uma MLP-ANN para a impedir o superajuste da rede. Portanto, o conjunto treinamento da MLP-ANN se restringe ao conjunto treinamento dos estudos de PLS, com exceção dos seis compostos selecionados para o conjunto validação.

O treinamento das MLP-ANNs foi feito variando-se as funções de transferência utilizadas e o número de neurônios na camada intermediária (de 1 a 20 neurônios). Foram utilizadas as três funções de transferência mais empregadas no treinamento de MLP-ANNs: *logsig*, *tansig* e *purelin*, combinadas das seguintes formas: *logsig-logsig*, *tansig-tansig*, *purelin-purelin*, *logsig-tansig*, *logsig-purelin*, *tansig-logsig*, *tansig-purelin*, *purelin-logsig*, *purelin-tansig*. Os resultados obtidos foram avaliados com base nos valores de r^2 , EQM e DMA. Primeiramente, foi escolhida a combinação de funções de transferência que resultasse em modelos de MPL-ANN com os menores valores de EQM e DMA e o maiores valores de r^2 . Uma vez selecionadas as funções de transferência, o número ideal de neurônios na camada intermediária também foi definido analisando-se os mesmos parâmetros estatísticos mencionados. Foram observados também os valores de $r^2_{\text{treinamento}}$, r^2_{teste} e $r^2_{\text{validação}}$ para a escolha da MLP-ANN com melhor desempenho.

Após uma análise prévia dos parâmetros de treinamento, as MLP-ANNs foram treinadas utilizando-se a função de treinamento *traingdm* que atualiza os valores dos pesos e *bias* de acordo com um gradiente descendente com *momentum*. Para o treinamento definiu-se uma taxa de aprendizado de 0,5 e constante de *momentum* de 0,9.

O treinamento das MLP-ANNs é interrompido se a magnitude do gradiente é menor que o valor pré-estabelecido de 1×10^{-5} ou se o número de verificações da validação aumentar mais que seis vezes desde a última vez que ele diminui. Essa última condição é utilizada para evitar um superajuste da rede, pois quando ele se inicia, o valor de EQM da validação começa a aumentar. Também definiu-se um número máximo de iterações de 100000. Caso esse valor seja atingido, a rede também interrompe o processo de treinamento.

4 RESULTADOS E DISCUSSÃO

4.1 Seleção de descritores

A partir dos cálculos de otimização de geometria e de frequências vibracionais dos compostos estudados, foram obtidos 14 descritores eletrônicos tais como: energia da molécula, polarizabilidade, momento de dipolo e energia dos orbitais de fronteira. Também foram obtidos 1205 descritores moleculares distribuídos em diferentes classes como, por exemplo, descritores topológicos, constitucionais, geométricos e índices de conectividade.

Uma pré-seleção dos descritores moleculares foi feita analisando-se os valores de r para cada descritor em relação aos valores de pK_i . Os descritores que apresentaram baixa correlação (valores de r menores que 0,3, em módulo) foram excluídos, restando ao todo 218 descritores.

A técnica de algoritmo genético permitiu selecionar dentre os descritores restantes aqueles de maior relevância para a construção de modelos matemáticos. Para o conjunto de compostos estudados, observou-se que seis descritores correspondem a um número ótimo, uma vez que uma quantidade menor de descritores pode gerar um modelo com baixa capacidade de predição enquanto que um número elevado dificulta ou até mesmo impossibilita a interpretação química do mesmo.

Aplicando-se a técnica de seleção de variáveis por algoritmo genético foram gerados 105 modelos, variando-se o número de gerações e o critério de corte para a correlação entre os descritores. Os melhores modelos tiveram seus descritores submetidos a análises prévias de PLS. A partir dessas análises identificou-se que o melhor modelo gerado foi aquele que continha os descritores GGI6 (índice de carga topológica de Gálvez de ordem 6), Mor23m (sinal 23 dos descritores 3D-MoRSE ponderado pela massa atômica), SRW06 (contagem de caminho de auto-retorno de ordem 6), H7m (autocorrelação H de comprimento 7 ponderada pela massa atômica), MLOGP (coeficiente de partição octanol-água de Moriguchi) e μ (momento de dipolo) (Tabela 2). Os valores dos descritores selecionados para cada um dos compostos estão dispostos na Tabela 3.

Tabela 2 - Parâmetros estatísticos do melhor modelo gerado a partir da técnica de seleção de variáveis por algoritmo genético.

Descritores: GGI6, Mor23m, SRW06, H7m, MLOGP e μ .	r^1	s^2	F^3	q^2 ⁴
	0,859	0,308	26,813	0,656

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

Tabela 3 - Valores dos descritores selecionados para os antagonistas do Sig-1R (os compostos do conjunto teste estão sublinhados).

Composto	GGI6	Mor23m	SRW06	H7m	MLOGP	μ (D)
<u>1</u>	0,379	-0,041	824	0,005	2,740	3,9145
<u>2</u>	0,344	-0,104	792	0,006	3,527	3,6026
3	0,379	-0,097	824	0,007	3,760	3,5344
4	0,374	-0,064	832	0,009	3,987	3,3577
5	0,420	-0,153	874	0,009	3,987	3,4229
6	0,564	-0,164	1266	0,023	3,171	2,9645
7	0,420	-0,084	874	0,008	2,972	3,9243
8	0,521	-0,130	996	0,038	3,395	4,1353
9	0,539	-0,109	1078	0,028	4,099	3,6387
10	0,400	-0,131	748	0,010	3,527	3,2825
11	0,460	-0,196	844	0,020	3,987	3,3283
<u>12</u>	0,344	-0,495	792	0,007	2,710	6,7072
13	0,283	-0,072	768	0,002	2,502	3,6585
14	0,249	-0,189	736	0,002	3,290	3,5950
15	0,283	-0,149	768	0,003	3,527	3,5286
16	0,278	-0,180	776	0,006	3,760	3,3482
<u>17</u>	0,324	-0,167	818	0,004	3,760	3,4088
18	0,365	-0,147	894	0,011	2,891	4,6254
19	0,542	-0,106	900	0,090	3,199	2,7169
20	0,441	-0,347	1028	0,063	3,620	4,0213
21	0,460	0,069	844	0,027	2,972	3,3173
<u>22</u>	0,456	-0,038	812	0,014	3,760	2,5741
23	0,501	-0,024	894	0,019	3,199	2,9389
24	0,339	-0,048	788	0,018	2,740	2,6634
25	0,339	-0,099	788	0,009	3,760	2,5825
26	0,380	-0,075	838	0,012	2,972	2,8193
27	0,379	-0,136	864	0,025	3,199	4,3134
28	0,374	-0,144	832	0,040	3,987	2,7419
29	0,379	-0,120	864	0,026	4,210	3,0999
30	0,539	-0,377	1118	0,062	5,232	3,3899
31	0,420	-0,034	914	0,034	3,422	2,8863
32	0,400	-0,211	788	0,026	3,987	3,2956
33	0,374	-0,541	832	0,020	3,180	6,8456
34	0,257	-0,101	808	0,014	2,972	4,1305
35	0,257	-0,279	808	0,012	3,760	4,6612
<u>36</u>	0,339	-0,097	908	0,019	3,422	4,2308
<u>37</u>	0,278	-0,176	732	0,015	3,760	3,2118
38	0,338	-0,268	898	0,024	4,407	3,3507

Composto	GGI6	Mor23m	SRW06	H7m	MLOGP	μ (D)
39	0,278	-0,082	746	0,016	2,972	2,7398
40	0,298	0,059	828	0,024	3,199	3,3934
41	0,545	0,103	984	0,147	1,751	1,6431
42	0,434	-0,161	838	0,011	1,968	4,3105
43	0,516	-0,116	888	0,014	2,200	3,3073
44	0,522	-0,030	886	0,006	2,972	4,2618
45	0,562	-0,158	962	0,017	2,346	4,7528
46	0,273	-0,224	768	0,006	2,234	3,7246
47	0,187	-0,273	712	0,003	1,991	2,2494
48	0,323	-0,216	824	0,041	2,740	2,3869
49	0,339	-0,399	932	0,006	2,465	2,7586
50	0,202	-0,428	876	0,002	2,238	1,9673
51	0,482	-0,387	994	0,004	2,688	2,6120
52	0,122	-0,209	662	0,004	1,471	2,9011
53	0,339	-0,397	932	0,007	3,476	1,0742
54	0,304	-0,418	900	0,006	3,253	1,1464
55	0,360	-0,465	856	0,005	3,645	1,1667
56	0,339	-0,353	1032	0,010	2,907	2,4861
57	0,420	-0,430	1058	0,022	2,834	5,4842
58	0,339	-0,413	972	0,021	2,907	2,5744
59	0,420	-0,480	1072	0,029	3,333	2,4399
60	0,420	-0,439	1098	0,042	3,260	5,5095
61	0,379	-0,089	824	0,004	3,121	3,8210
62	0,405	-0,176	818	0,027	3,431	7,1464
63	0,283	-0,104	768	0,002	2,502	5,1964
64	0,283	-0,137	768	0,003	3,527	4,7842

4.2 Análises de PLS

Os estudos de PLS permitiram propor um modelo capaz de prever satisfatoriamente os valores de pK_i dos compostos investigados. O modelo PLS foi obtido utilizando-se o conjunto treinamento contendo 51 compostos, validado externamente com o conjunto teste (13 compostos). Os resultados das análises de PLS obtidos para o conjunto treinamento apontam para um número ótimo de VLs igual a 5, descrevendo 82% da informação original dos dados. A Tabela 4 apresenta os modelos obtidos, bem como os parâmetros estatísticos utilizados para a escolha do melhor modelo.

Tabela 4 - Variância percentual acumulada, SEV, SEC, PRESS, q^2 e r^2 dos modelos PLS.

<i>VL</i>	<i>%Var_{acumulada}</i> ¹	<i>SEV</i> ²	<i>PRESS_{val}</i> ³	<i>SEC</i> ⁴	<i>PRESS_{cal}</i> ⁵	<i>q²</i> ⁶	<i>r²</i> ⁷
1	23,49	0,406	8,425	0,360	6,362	0,445	0,581
2	46,69	0,350	6,248	0,294	4,149	0,588	0,726
3	56,54	0,312	4,953	0,280	3,679	0,673	0,757
4	68,33	0,313	5,008	0,279	3,578	0,670	0,764
5	82,27	0,307	4,792	0,279	3,514	0,684	0,768
6	100,00	0,307	4,797	0,283	3,512	0,684	0,768

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.

A escolha do modelo PLS obtido a partir de cinco VLs é justificada por este modelo apresentar os menores valores de SEV (0,307) e SEC (0,279) e o menor valor de PRESS_{val} (4,792) e o segundo menor valor de PRESS_{cal} (3,514). Os valores de q^2 e r^2 obtidos com cinco VLs são os melhores valores encontrados para as diferentes VLs. O valor de q^2 foi de 0,684 enquanto que o valor de r^2 foi de 0,768. Os valores desses coeficientes permitem inferir que o modelo PLS obtido é um bom modelo de regressão, pois, de acordo com a literatura, o mesmo deve apresentar $q^2 > 0,5$ e $r^2 > 0,6$.^{58,59}

É possível observar que à medida que se aumenta o número de VLs, há um aumento no valor de r^2 , uma vez que mais informação vai sendo incluída no modelo. O q^2 também tem seus valores aumentados com um número maior de VLs. Neste caso, observa-se que com cinco e seis VLs tem-se o mesmo valor de r^2 e q^2 . Entretanto, a escolha do melhor modelo deve basear-se principalmente nos valores de SEV e SEC. O menor valor de SEC foi observado para os modelos com 4 e 5 VLs e o menor valor de SEV, para os modelos com 5 e 6 VLs. De forma geral, o modelo com 5 VLs se apresenta como o melhor modelo PLS para esse caso.

A Figura 6 traz a contribuição de cada descritor para o vetor de regressão do modelo PLS (GGI6: -0,283; Mor23m: 0,591; SRW06: 0,184; H7m: -0,575; MLOGP: 0,488; μ : -0,249). É possível observar que os descritores Mor23m, H7m e MLOGP têm uma maior contribuição para o vetor de regressão, sendo portanto, mais relevantes para descrever a atividade biológica dos compostos investigados.

Figura 6 - Coeficientes de regressão para o modelo PLS.

A análise de *outliers* se faz necessária para a detecção de amostras atípicas que devem ser analisadas na obtenção dos modelos matemáticos. As análises do gráfico de *leverage* versus resíduos de Student para o modelo PLS mostram que os compostos **6** e **41** apresentaram valores altos de *leverage*, mas com valores de resíduos de Student dentro do limite aceitável (Figura 7). Já os compostos **40** e **43** estão fora dos limites ideais para os resíduos de Student, porém dentro do limite de *leverage*.

Figura 7 - Detecção de *outliers* no modelo PLS.

É importante notar que o composto **6** apresenta um substituinte mais volumoso na posição R2 em relação aos demais compostos e o composto **41** possui um grupo com átomo de enxofre na posição X, sendo o único composto com esse substituinte (Figura 8). Os compostos **40** e **43** também possuem um grupo substituinte na posição X que não é comum aos demais compostos estudados, com exceção do composto **42**. Tais características estruturais dos compostos mencionados podem ter contribuído para que os mesmos se destacassem nas análises de *outliers*.

Figura 8 - Compostos identificados na análise de *outliers* com valores altos de *leverage* (compostos **6** e **41**) e valores fora dos limites dos resíduos de Student (compostos **40** e **43**). Em destaque, os grupos substituintes que se diferenciam dos demais compostos do conjunto treinamento.

Novos modelos de PLS foram gerados removendo-se cada um dos compostos **6**, **40**, **41** ou **43**, porém não foram observadas mudanças significativas nesses modelos. Por isso, o conjunto inicial incluindo todos os compostos foi mantido.

Após a escolha do melhor modelo, o mesmo deve ser validado ainda quanto à sua robustez e à presença de correlações ao acaso. Para isso foram realizados testes de validação cruzada com a exclusão de N compostos (*leave-N-out*) e testes de randomização da atividade biológica.

Para realizar os testes de *leave-N-out* foram excluídos de 1 a 10 compostos, sendo feitas 30 repetições para cada valor de N excluído (Figura 9). O modelo PLS se mostra estável uma vez que os valores de q^2 são altos e os valores de desvio para cada valor de N são menores que 0,030. Para um modelo ser considerado robusto é necessário que os valores de q^2 sejam próximos dos valores do teste de *leave-one-out* e os desvios devem ser menores que 0,1.⁴⁸ Portanto, diante dos resultados obtidos pode-se inferir que o modelo PLS é estável e robusto.

Figura 9 - Testes de *leave-N-out* aplicados ao modelo PLS.

Os testes de randomização da atividade biológica demonstram que os valores de q^2 e r^2 são menores que -0,038 e 0,278, respectivamente, com exceção do modelo original (Figura 10). Além disso, os valores dos interceptos nos gráficos de $|r|$ versus q^2 e $|r|$ versus r^2 foram, respectivamente, -0,305 e 0,051. Um modelo construído sem correlações ao acaso deve apresentar valores de q^2 e r^2 menores que 0,3 a 0,4 para os modelos randomizados e os interceptos dos gráficos de $|r|$ versus q^2 e $|r|$ versus r^2 devem ser menores que 0,05 e 0,3, respectivamente.^{60,61} Portanto, é possível afirmar que o modelo A não foi obtido a partir de correlações ao acaso.

Figura 10 - Testes de randomização dos valores de pK_i aplicados ao modelo PLS.

A validação externa do modelo PLS foi feita utilizando-se o conjunto teste previamente escolhido. A predição dos valores de pK_i do conjunto teste apresentou $r^2_{\text{pred}} = 0,785$, com $\text{SEV} = 0,325$ e $\text{PRESS} = 1,371$. Os valores de pK_i experimental e predito dos conjuntos treinamento e teste estão apresentados na Figura 11.

Figura 11 - Predição da atividade biológica pelo modelo PLS.

Os resultados de predição para o conjunto teste mostram uma boa capacidade preditiva do modelo PLS, uma vez que os resíduos (erros de predição) são menores que 10% com exceção do composto **45** (13,84%), sendo o valor médio dos resíduos relativos de apenas 3,2% (Tabela 5). Analisando-se a estrutura química do composto **45** é possível observar que ele possui um substituinte $-\text{COCH}_3$ na posição R_3 , que não está presente nos demais compostos do conjunto treinamento. Possivelmente, a predição da atividade biológica desse composto apresentou um resíduo mais elevado em função da presença desse substituinte.

Tabela 5 - Resíduos de predição do conjunto teste feita pelo modelo PLS.

<i>Composto</i>	<i>pK_i experimental</i>	<i>pK_i predito</i>	<i>Resíduos Absolutos</i>	<i>Resíduos Relativos (%)*</i>
1	7,967	7,814	0,152	1,91
2	8,284	8,054	0,230	2,77
12	6,532	6,533	-0,001	0,02
17	8,066	8,121	-0,056	0,69

<i>Composto</i>	<i>pK_i experimental</i>	<i>pK_i predito</i>	<i>Resíduos Absolutos</i>	<i>Resíduos Relativos (%)*</i>
22	7,983	8,135	-0,152	1,91
36	7,830	7,899	-0,069	0,89
37	7,764	7,986	-0,221	2,85
45	6,130	6,978	-0,848	13,84
46	6,745	7,364	-0,620	9,19
48	7,357	7,262	0,094	1,28
52	7,636	7,366	0,270	3,54
53	8,000	7,837	0,163	2,04
64	8,022	7,966	0,056	0,70

*Os resíduos relativos são obtidos dividindo-se os valores dos resíduos absolutos (em módulo) pelos valores de pK_i experimental correspondente e multiplicando-se o resultado por 100.

A capacidade preditiva do modelo PLS deve ser avaliada não só a partir de valores de q^2 , mas também analisando-se os parâmetros propostos por Golbraikh e Tropsha (2002) e Zhang e colaboradores (2006).^{58,62} Segundo os autores, um modelo com capacidade preditiva deve obedecer pelo menos uma das duas condições seguintes:

$$1) \ r^2_{pred} > 0,7;$$

$$\frac{(r^2_{pred} - r_0^2)}{r^2_{pred}} < 0,1 \quad e \quad 0,85 < k < 1,15$$

$$2) \ (r_0^2 - r_0'^2) < 0,3;$$

$$\frac{(r^2_{pred} - r_0'^2)}{r^2_{pred}} < 0,1 \quad e \quad 0,85 < k' < 1,15$$

onde r_0^2 e $r_0'^2$ são coeficientes de determinação (pK_i predito *versus* pK_i experimental e pK_i experimental *versus* pK_i predito, respectivamente) e k e k' são as inclinações das retas de regressão que passam pela origem.

O modelo PLS tem sua capacidade preditiva confirmada uma vez que ele satisfaz a primeira condição: $r^2_{pred} = 0,785$, $(r^2_{pred} - r_0^2)/r^2_{pred} = 0,031$ e $k = 0,991$. Os parâmetros estatísticos obtidos para o conjunto teste estão dispostos na Tabela 6.

Tabela 6 - Parâmetros estatísticos para a avaliação da capacidade preditiva do modelo PLS.

	r_{pred}^2	r_0^2	$r_0'^2$	k	k'	$\frac{(r_{pred}^2 - r_0^2)}{r_{pred}^2}$	$\frac{(r_{pred}^2 - r_0'^2)}{r_{pred}^2}$	$(r_0^2 - r_0'^2)$
Modelo PLS	0,785	0,761	0,553	0,991	1,008	0,031	0,296	0,208

O modelo PLS possui valores aceitáveis para os parâmetros estatísticos, além de apresentar resultados satisfatórios para os testes de validação. A análise do gráfico de *leverage versus* resíduos de Student não detectou compostos atípicos, logo todos os compostos foram mantidos na construção do modelo. Além disso, os testes de *leave-N-out* e de randomização da atividade biológica mostraram a robustez do modelo bem como a inexistência de correlações ao acaso. Após ser devidamente validado, o modelo PLS teve sua capacidade preditiva comprovada a partir da validação externa utilizando o conjunto teste. Portanto, o modelo PLS obtido se mostra uma ferramenta útil para a predição da atividade biológica de novos antagonistas do Sig-1R pertencentes à classe estudada neste trabalho.

4.3 Análises de ANN

Com o propósito de investigar se os descritores selecionados apresentam alguma relação não linear com a atividade biológica, foram obtidos modelos de ANN treinando-se redes de alimentação para frente do tipo MLP-ANN para prever os valores de pK_i . Inicialmente, foram feitos testes para definir o número ótimo de neurônios na camada intermediária, bem como a função de transferência mais adequada. A Figura 12 apresenta a distribuição do EQM na predição dos valores de pK_i para as diferentes MLP-ANNs. Nesta análise, variou-se o número de neurônios na camada intermediária de 1 a 20 e também investigou-se o comportamento das MLP-ANNs frente a combinação de duas a duas das três funções de transferência: *logsig*, *tansig* e *purelin*. Tal combinação é possível pois as funções de transferência atuam na camada intermediária e na camada de saída durante o treinamento da MLP-ANN. No apêndice A encontram-se os valores de r^2 , EQM e DMA para todas as MLP-ANNs treinadas.

Figura 12 - Desempenho das MLP-ANNs em relação ao número de neurônios na camada intermediária e às funções de transferência utilizadas.

Analisando-se a Figura 12 e o apêndice A é possível perceber que as MLP-ANNs treinadas utilizando-se as funções de transferência *tansig-tansig*, *tansig-purelin*, *logsig-tansig* e *logsig-purelin* apresentaram os menores valores de EQM e DMA e os maiores valores de r^2 . Esses resultados mostram que a escolha das funções de transferência no treinamento de MLP-ANNs é fundamental, uma vez que elas podem produzir resultados diferentes entre si.

Para cada conjunto de funções de transferência utilizado, determinou-se o número ótimo de neurônios na camada intermediária capaz de produzir o modelo com os melhores valores de EQM, r^2 e DMA. Além disso, avaliou-se os valores de r^2 para os conjuntos treinamento, validação e teste, utilizados simultaneamente no treinamento das MLP-ANNs. A Tabela 7 contém os melhores modelos gerados para cada conjunto de funções de transferência.

Tabela 7 - Melhores modelos de MLP-ANN gerados a partir de diferentes combinações das funções de transferência *tansig*, *logsig* e *purelin*.

<i>Funções de transferência</i>	<i>Arquitetura</i>	<i>EQM</i>	<i>r²</i>	<i>DMA</i>	<i>r² treinamento</i>	<i>r² validação</i>	<i>r² teste</i>	<i>Número de iterações</i>
<i>tansig-tansig</i>	6-12-1	0,0070	0,8309	0,0626	0,8907	0,8520	0,7929	3789
<i>logsig-logsig</i>	6-10-1	0,0150	0,6400	0,0739	0,8441	0,9072	0,6323	100000
<i>purelin-purelin</i>	6-18-1	0,0101	0,7582	0,0767	0,7493	0,9373	0,7929	36

<i>Funções de transferência</i>	<i>Arquitetura</i>	<i>EQM</i>	r^2	<i>DMA</i>	r^2 <i>treinamento</i>	r^2 <i>validação</i>	r^2 <i>teste</i>	<i>Número de iterações</i>
<i>tansig-logsig</i>	6-13-1	0,0138	0,6692	0,0651	0,8750	0,7374	0,5137	100000
<i>tansig-purelin</i>	6-13-1	0,0063	0,8493	0,0623	0,8811	0,8322	0,8091	774
<i>logsig-tansig</i>	6-9-1	0,0073	0,8239	0,0636	0,8694	0,7867	0,8091	5018
<i>logsig-purelin</i>	6-20-1	0,0069	0,8331	0,0616	0,8997	0,8639	0,6964	5831
<i>purelin-tansig</i>	6-12-1	0,0085	0,7960	0,0720	0,7937	0,8636	0,8214	71
<i>purelin-logsig</i>	6-10-1	0,0188	0,5477	0,1009	0,6539	0,4951	0,6760	910

Avaliando-se os resultados obtidos escolheu-se a MLP-ANN gerada utilizando-se as funções de transferência *tansig-tansig* e 12 neurônios na camada intermediária como sendo o modelo de MLP-ANN de melhor desempenho obtido nesse estudo. Além de apresentar alguns parâmetros estatísticos superiores aos modelos gerados utilizando as funções de transferência *tansig-purelin*, *logsig-tansig* e *logsig-purelin*, os modelos obtidos com a função de transferência *tansig-tansig* se mostraram mais estáveis durante o processo de treinamento das MLP-ANNs.

A arquitetura da MLP-ANN selecionada é 6-12-1 (seis neurônios na camada de entrada, 12 na camada intermediária e 1 na camada de saída). Na Figura 13 é apresentado um diagrama esquemático do modelo MLP-ANN selecionado.

Figura 13 - Diagrama esquemático da MLP-ANN selecionada.

A MLP-ANN selecionada possui 12x6 pesos e 12 valores de *bias* que conectam a camada de entrada e a camada intermediária e 1x12 pesos e 1x1 valores de *bias* conectando a camada intermediária e a camada de saída. Os valores correspondentes estão dispostos na Tabela 8.

Tabela 8 - Valores de pesos e *bias* do modelo de MLP-ANN selecionado.

Pesos conectando a camada de entrada e a camada intermediária ($w_{12 \times 6}$)											
0,8441	-0,4459	0,7403	-0,9682	-1,3340	0,5004						
1,0058	-0,0135	0,8431	1,0948	-0,4443	0,9382						
-0,7092	-0,5882	0,6771	-0,9142	1,2066	-0,6971						
0,2837	-1,3777	-1,1258	-0,4781	-0,7768	0,0011						
0,8327	1,0063	-0,5471	-0,8511	-1,5322	0,6034						
1,5349	-1,0109	-0,8832	0,8859	-0,6669	0,1580						
1,1872	-1,3776	-0,8546	0,5453	-1,2266	0,8714						
0,2607	-1,2028	-1,3965	0,0979	0,9930	0,8817						
0,7988	-1,0566	1,4090	-0,4374	1,1587	0,0087						
0,1880	-0,2352	-0,9419	0,3603	1,6486	0,1333						
-0,5458	0,2826	-0,2831	1,3868	-0,1187	1,5098						
0,6023	-0,3217	-1,1768	-0,8229	-0,0213	-1,4176						
Bias na camada intermediária ($b_{12 \times 1}$)											
-2,0690	-1,8563	1,4403	-1,1082	-0,4854	-0,5422	-0,5801	0,7106	0,8931	-1,6895	-1,5192	2,0555
Pesos conectando a camada intermediária e a camada de saída ($w_{1 \times 12}$)											
0,2379	0,3561	0,3134	0,0436	0,5540	-0,3565	-0,8385	0,5051	0,2771	0,4694	-0,3582	0,1441
Bias na camada de saída ($b_{1 \times 1}$)											
-0,5109											

Os valores de EQM para o modelo MLP-ANN selecionado é de 0,0070, sendo o $r^2 = 0,8309$ e $DMA = 0,0626$. A MLP-ANN foi treinada até a iteração 3789, conforme ilustrado na Figura 14. Como esperado para uma MLP-ANN onde não haja um superajuste do modelo, os valores de EQM para o conjunto treinamento são menores do que para os conjuntos validação e teste.⁵⁰ Comportamentos semelhante foi observado para as demais MLP-ANNs obtidas nesse estudo.

Figura 14 - EQM para os conjuntos treinamento, validação e teste em função do número de iterações.

O modelo MLP-ANN apresenta ainda valores de r^2 para os conjuntos treinamento, validação e teste de 0,8907, 0,8520 e 0,7929, respectivamente, além de baixos valores para os erros de predição da atividade biológica do conjunto teste. As etapas de calibração, validação e teste são feitas simultaneamente durante o processo de treinamento da MLP-ANN. A Figura 15 apresenta a correlação entre os valores experimentais de pK_i e os valores preditos pelo modelo de MLP-ANN.

Figura 15 - Predição da atividade biológica pelo modelo MLP-ANN.

A predição dos valores de pK_i para o conjunto teste realizada pelo modelo MLP-ANN apresenta erros de predição menores que 10% com exceção dos compostos **45** (11,1%) e **46** (12,7%) e um valor médio dos resíduos de 3,34% (Tabela 9). Vale ressaltar que o composto **45** também apresentou resíduo maior que 10% na predição da atividade biológica utilizando o modelo PLS, conforme discutido anteriormente. Portanto, o modelo MLP-ANN também apresenta boa capacidade preditiva.

Tabela 9 - Resíduos de predição do conjunto teste feita pelo modelo MLP-ANN selecionado.

<i>Composto</i>	<i>pK_i experimental</i>	<i>pK_i predito</i>	<i>Resíduos Absolutos</i>	<i>Resíduos Relativos (%)*</i>
1	7,967	7,854	0,112	1,41
2	8,284	8,058	0,226	2,73
12	6,532	6,725	-0,194	2,97
17	8,066	8,039	0,027	0,33
22	7,983	8,046	-0,063	0,80
36	7,830	8,004	-0,174	2,23
37	7,764	7,944	-0,179	2,31
45	6,130	6,808	-0,678	11,05
46	6,745	7,603	-0,858	12,73
48	7,357	7,435	-0,078	1,07
52	7,636	7,620	0,016	0,21
53	8,000	7,586	0,414	5,18
64	8,022	8,057	-0,034	0,43

*Os resíduos relativos são obtidos dividindo-se os valores dos resíduos absolutos (em módulo) pelos valores de pK_i experimental correspondente e multiplicando-se o resultado por 100.

4.4 Proposição de modelos QSAR

Diante dos modelos obtidos pelas análises de PLS e ANN, propõe-se dois modelos QSAR para a predição da atividade biológica de compostos da classe de 1-arilpirazóis, antagonistas do Sig-1R. O primeiro, o modelo gerado a partir de análises de PLS, foi capaz de descrever relações lineares entre os descritores selecionados neste estudo e a atividade biológica dos compostos, apresentando $r^2 = 0,768$, $q^2 = 0,684$ e $r^2_{\text{teste}} = 0,785$. Este modelo foi devidamente validado de forma a demonstrar sua robustez e alta capacidade preditiva. Obteve-se ainda um segundo modelo QSAR, obtido a partir de análises de MLP-ANN, com $r^2_{\text{treinamento}} = 0,891$, $r^2_{\text{validação}} = 0,852$ e $r^2_{\text{teste}} = 0,793$. Este modelo descreve as relações não

lineares entre os descritores e os valores de atividade biológica dos compostos, tão comum para sistemas biológicos.

4.5 Descritores selecionados e a interpretação química dos modelos obtidos

Os descritores selecionados a partir da técnica de algoritmo genético compreendem cinco descritores moleculares (GGI6, Mor23m, SRW06, H7m e MLOGP) e um descritor eletrônico (μ). Tais descritores foram utilizados nos estudos de PLS e ANN para a geração de modelos de predição de pK_i dos compostos estudados. Portanto, torna-se imprescindível compreender o significado de cada descritor, a fim de que uma interpretação química dos modelos possa ser desenvolvida.

Os estudos envolvendo ligantes do Sig-1R identificaram uma região importante no centro das moléculas no qual estão presentes pequenos grupos contendo átomos de nitrogênio.⁶³ Estudos posteriores mostram que átomos de nitrogênio em anéis de piperidina localizados nessa região são capazes de estabelecerem interações com o resíduo Asp126 do Sig-1R.¹⁹ Tal interação é fundamental para a atividade biológica dos antagonistas do Sig-1R e indica que a distribuição da densidade eletrônica na molécula possa ser um requisito essencial para o planejamento de novos ligantes. Dentre os descritores selecionados pelos estudos de QSAR, o GGI6, associado à transferência de carga na molécula, fornece indícios de que a distribuição da carga na molécula é fundamental para a sua atividade biológica. Além disso, o descritor μ também foi selecionado nesse estudo, contribuindo assim para evidenciar que a densidade eletrônica nos antagonistas do Sig-1R tem relação com sua interação com os resíduos do sítio de ligação do receptor.

O descritor GGI6, índice de carga topológica de ordem 6, pertence ao grupo de descritores topológicos e está associado a índices de carga. Os índices de carga topológica de Gálvez são capazes de indicar a transferência de carga em uma molécula, a partir da transferência de carga entre pares de átomos.^{64,65} Já o descritor μ está associado à distribuição de carga e é obtido a partir da primeira derivada da energia em relação a um campo elétrico aplicado.⁶⁶ Esse descritor tem se mostrado muito importante para a compreensão de interações entre ligantes e receptores.^{67,68} De forma geral, os compostos estudados que apresentam maiores valores de atividade biológica possuem valores mais baixos tanto para o descritor GGI6 quanto para o descritor μ em comparação com os compostos menos ativos.

Essa análise sugere que os compostos com grupos substituintes menos polares são mais eficazes para a interação com o Sig-1R

Além da presença de átomos de nitrogênio nos ligantes do Sig-1R, os pesquisadores mostram ainda que são necessárias duas regiões hidrofóbicas a determinada distância dos grupos contendo átomos de nitrogênio. O tamanho dos substituintes nessas regiões hidrofóbicas está associado também à atividade biológica dos compostos.^{19,63} A seleção dos descritores SRW06, Mor23m e H7m evidencia a distribuição dos átomos ao longo da estrutura dos compostos como um fator fundamental para a predição da atividade biológica. Além disso, os descritores Mor23m e H7m ainda sugerem que a distribuição da massa dos compostos ao longo de sua estrutura está associada à atividade biológica. Em geral, os compostos mais ativos apresentaram valores mais altos para o descritor Mor23m e valores mais baixos para os descritores SRW06 e H7m quando comparados aos compostos menos ativos.

O descritor SRW06 é um descritor topológico obtido a partir da contagem de trilhas em um grafo molecular (objeto geométrico para representar uma molécula onde os átomos constituem os vértices e as ligações químicas, as arestas). As trilhas são constituídas de qualquer sequência de arestas adjacentes em um grafo. Para determinar o valor desse descritor, deve-se inicialmente determinar a contagem de trilhas atômicas.^{69,70} O descritor SRW06 é um descritor de contagem de caminho de retorno de comprimento igual a 6.

O descritor Mor23m pertence à classe dos descritores 3D-MoRSE (do inglês, *3D Molecule Representation of Structures based on Electron diffraction*), que se baseiam em métodos que utilizam dados de difração eletrônica para gerar curvas de espalhamento teóricas.⁷¹⁻⁷³ Os descritores 3D-MoRSE são representados por $Morsw$, onde s representa a direção de espalhamento e pode assumir valores no intervalo $1 \leq s \leq 32$ e w representa as diferentes propriedades atômicas dos átomos i e j . Para o descritor Mor23m, o valor de s é 23, enquanto que a massa atômica é a propriedade considerada no cálculo desse descritor.

O descritor H7m pertence à classe de descritores GETAWAY (do inglês, *Geometry, Topology and Atom-Weights Assembly*), que codifica informações geométricas, topológicas e químicas. Dentro dessa classe, encontram-se os descritores H -GETAWAY, obtidos a partir da matriz de influência molecular (H). O descritor H7m é classificado como de autocorrelação, sendo que o índice H permite inferir que para o átomo i , somente átomos j a uma distância topológica d_{ij} têm possibilidade de interagir com o átomo i .^{74,75} Para o cálculo do descritor

H7m, leva-se em consideração a massa dos átomos, enquanto que se define uma distância topológica igual a 7.

O descritor MLOGP refere-se ao cálculo do coeficiente de partição octanol-água de Moriguchi. O modelo utilizado para prever o coeficiente de partição foi desenvolvido a partir de um estudo de relação estrutura-logP, utilizando análises de regressão múltipla aplicadas a 1230 moléculas orgânicas.⁷⁶ O coeficiente de partição octanol-água representa a movimentação de um composto entre uma fase orgânica e uma fase aquosa, sendo, portanto, capaz de descrever a capacidade do composto em penetrar na membrana celular. Tal coeficiente é utilizado como uma medida de lipofilicidade, que está associada à permeabilidade, biodisposição e toxicidade de um determinado composto. Os fármacos que apresentam valores altos de coeficiente de partição geralmente são encontrados nas bicamadas lipídicas das células enquanto que aqueles que têm valores baixos do mesmo coeficiente podem ser encontrados em regiões hidrofílicas como o soro sanguíneo.⁷⁷ Uma vez que o receptor Sig-1R é uma proteína transmembranar ancorada no retículo endoplasmático e em membranas plasmáticas^{16,17}, espera-se que os compostos capazes de se ligar a ele possuam um valor alto de coeficiente de partição. Os estudos aqui desenvolvidos identificaram que os compostos mais ativos apresentam, de forma geral, valores mais altos de MLOGP em comparação aos compostos com menores valores de atividade biológica.

Diante do exposto, é possível inferir que as informações contidas nos descritores selecionados são capazes de descrever a atividade biológica dos compostos investigados. Dessa forma, torna-se importante utilizar os modelos QSAR propostos neste estudo para a predição de valores de atividade biológica de novos compostos da classe de 1-arilpirazóis, contribuindo assim para o planejamento de novos fármacos para o tratamento da DN.

5 CONCLUSÕES E PERSPECTIVAS

O estudo QSAR desenvolvido a partir de análise de PLS e ANN permitiu propor dois modelos matemáticos capazes de prever a atividade biológica de compostos da classe de 1-arilpirazóis, antagonistas do Sig-1R. Além disso, fatores importantes da estrutura dos compostos, associados à atividade biológica dos mesmos, foram identificados a partir dos descritores selecionados. Observou-se, de forma geral, que os compostos mais ativos apresentaram maiores valores de Mor23m e MLOGP e menores valores de GGI6, μ , SRW06 e H7m.

Os parâmetros estatísticos dos modelos gerados estão dentro dos limites aceitáveis e as validações realizadas comprovaram a capacidade preditiva dos mesmos. Validações adicionais aplicadas aos modelos PLS permitiram comprovar sua robustez e estabilidade, além de demonstrar que os mesmos não foram obtidos a partir de correlações ao acaso. As análises de MLP-ANN permitiram identificar a arquitetura 6-12-1, treinada com as funções de transferência *tansig-tansig*, como a mais adequada para descrever a atividade biológica dos compostos investigados, utilizando os descritores selecionados.

A proposição de dois modelos QSAR distintos permite que tanto relações lineares quanto não lineares entre o conjunto de descritores selecionado e a atividade biológica dos compostos estudados sejam descritas a partir de modelos matemáticos. Portanto, os dois modelos QSAR propostos se mostram promissores para a predição da atividade biológica de novos compostos da classe de 1-arilpirazóis, contribuindo significativamente para o planejamento de novos fármacos para o tratamento da DN.

Entretanto, espera-se desenvolver futuramente estudos de QSAR que utilizem outros tipos de descritores, tal como o HQSAR, de forma que sejam evidenciados novos fatores associados a atividade biológica dos compostos e que complementem os resultados já encontrados neste estudo. Dessa forma, será possível propor alterações estruturais que visem novos compostos da classe de 1-arilpirazóis com maior atividade biológica frente ao Sig-1R.

REFERÊNCIAS

1. 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.
2. SMITH, B. H.; TORRANCE, N. Epidemiology of neuropathic pain and its impact on quality of life. **Current Pain and Headache Reports**, v. 16, n.3, p. 191–198, 2012.
3. GALLUZZI, K. E. Managing neuropathic pain. **The Journal of the American Osteopathic Association**, v. 107, n. 11, p. ES39–ES48, 2007.
4. BOUREAU, F.; DOUBRÈRE, J. F.; LUU, M. Study of verbal description in neuropathic pain. **Pain**, v. 42, n. 2, p. 145–152, 1990.
5. TORRANCE, N.; SMITH, B. H.; BENNETT, M. I.; LEE, A. J. The epidemiology of chronic pain of predominantly neuropathic origin. Results from a general population survey. **The Journal of Pain**, v. 7, n. 4, p. 281–289, 2006.
6. BOUHASSIRA, D.; LANTÉRI-MINET, M.; ATTAL, N.; LAURENT, B.; TOUBOUL, C. Prevalence of chronic pain with neuropathic characteristics in the general population. **Pain**, v. 136, n. 3, p. 380–387, 2008.
7. BOWSHER, D. Neurogenic pain syndromes and their management. **British Medical Bulletin**, v. 47, n. 3, p. 644–666, 1991.
8. NIGHTINGALE, S. The neuropathic pain market. **Nature Reviews Drug Discovery**, v. 11, p. 101–102, 2012.
9. FREYNHAGEN, R.; BENNETT, M. I. Diagnosis and management of neuropathic pain. **BMJ**, v. 339:b3002, p. 391–395, 2009.
10. 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.; ALMANSA, C. Synthesis and biological evaluation of the 1-arylpyrazole class of σ_1 receptor antagonists: identification of 4-{2-[5-methyl-1-(naphthalen-2-yl)-1H-pyrazol-3-yloxy]ethyl}morpholine (S1RA, E-52862). **Journal of Medicinal Chemistry**, v. 55, n. 19, p. 8211–8224, 2012.

11. RIGANAS, S.; PAPANASTASIOU, I.; FOSCOLOS, G. B.; TSOTINIS, A.; BOURGUIGNON, J.-J.; SERIN, G.; MIRJOLET, J.-F.; DIMAS, K.; KOURAFALOS, V. N.; ELEUTHERIADES, A.; MOUTSOS, V. I.; KHAN, H.; GEORGAKOPOULOU, S.; ZANIOU, A.; PRASSA, M.; THEODOROPOULOU, M.; PONDIKI, S.; VAMVAKIDES, A. Synthesis, σ_1 , σ_2 -receptors binding affinity and antiproliferative action of new C1-substituted adamantanes. **Bioorganic & Medicinal Chemistry**, v. 20, n. 10, p. 3323–3331, 2012.

12. 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 sigma(1) receptor antagonist BD1047 blocks both mechanical allodynia and increases in spinal NR1 expression during the induction phase of rodent neuropathic pain. **Anesthesiology**, v. 109, n. 5, p. 879–889, 2008.

13. 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 chronic spinal dog. **The Journal of Pharmacology and Experimental Therapeutics**, v. 197, n. 3, p. 517–532, 1976.

14. TAM, S. W. Naloxone-inaccessible sigma receptor in rat central nervous system. **Proceedings of the National Academy of Sciences of the United States of America**, v. 80, n. 21, p. 6703–6707, 1983.

15. VAUPEL, D. B. Naltrexone fails to antagonize the sigma effects of PCP and SKF 10,047 in the dog. **European Journal of Pharmacology**, v. 92, n. 3-4, p. 269–274, 1983.

16. QUIRION, R.; BOWEN, W. D.; ITZHAK, Y.; JUNIEN, J. L.; MUSACCHIO, J. M.; ROTHMAN, R. B.; SU, T. P.; TAM, S. W.; TAYLOR, D. P. A proposal for the classification of sigma binding sites. **Trends in Pharmacological Sciences**, v. 13, n. 3, p. 85–86, 1992.

17. HELLEWELL, S. B.; BRUCE, A.; FEINSTEIN, G.; ORRINGER, J.; WILLIAMS, W.; BOWEN, W. D. Rat liver and kidney contain high densities of sigma 1 and sigma 2 receptors: characterization by ligand binding and photoaffinity labeling. **European Journal of Pharmacology**, v. 268, n. 1, p. 9–18, 1994.

18. KEKUDA, R.; PRASAD, P. D.; FEI, Y.-J.; LEIBACH, F. H.; GANAPATHY, V. Cloning and functional expression of the human type 1 sigma receptor (hSigmaR1). **Biochemical and Biophysical Research Communications**, v. 229, n. 2, p. 553–558, 1996.

19. LAURINI, E.; COL, V. D.; MAMOLO, M. G.; ZAMPIERI, D.; POSOCCO, P.; FERMEGLIA, M.; VIO, L.; PRICL, S. Homology model and docking-based virtual screening for ligands of the σ_1 receptor. **ACS Medicinal Chemistry Letters**, v. 2, n. 11, p. 834–839, 2011.

20. LAURINI, E.; MARSON, D.; COL, V. D.; FERMEGLIA, M.; MAMOLO, M. G.; ZAMPIERI, D.; VIO, L.; PRICL, S. Another brick in the wall. Validation of the σ_1 receptor 3D model by computer-assisted design, synthesis, and activity of new σ_1 ligands. **Molecular Pharmaceutics**, v. 9, n. 11, p. 3107–3126, 2012.
21. GELDENHUYS, W. J.; NOVOTNY, N.; MALAN, S. F.; VAN DER SCHYF, C. J. 3D-QSAR and docking studies of pentacycloundecylamines at the sigma-1 (σ_1) receptor. **Bioorganic & Medicinal Chemistry Letters**, v. 23, n. 6, p. 1707–1711, 2013.
22. AYDAR, E.; PALMER, C. P.; KLYACHKO, V. A.; JACKSON, M. B. The sigma receptor as a ligand-regulated auxiliary potassium channel subunit. **Neuron**, v. 34, n. 3, p. 399–410, 2002.
23. HAYASHI, T.; SU, T.-P. Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate Ca^{2+} signaling and cell survival. **Cell**, v. 131, n. 3, p. 596–610, 2007.
24. PAIN, R. H. **Mechanisms of protein folding**. 2. ed. New York: Oxford University, 2000. 433 p.
25. GUNDLACH, A. L.; LARGENT, B. L.; SNYDER, S. H. Autoradiographic localization of sigma receptor binding sites in guinea pig and rat central nervous system with (+)3H-3-(3-hydroxyphenyl)-N-(1-propyl)piperidine. **The Journal of Neuroscience**, v. 6, n. 6, p. 1757–1770, 1986.
26. MAURICE, T.; SU, T.-P. The pharmacology of sigma-1 receptors. **Pharmacology & Therapeutics**, v. 124, n. 2, p. 195–206, 2009.
27. WOLFE JR, S. A.; CULP, S. G.; DE SOUZA, E. B. Sigma-receptors in endocrine organs: identification, characterization, and autoradiographic localization in rat pituitary, adrenal, testis, and ovary. **Endocrinology**, v. 124, n. 3, p. 1160–1172, 1989.
28. HAYASHI, T.; SU, T.-P. Sigma-1 receptor ligands: potential in the treatment of neuropsychiatric disorders. **CNS Drugs**, v. 18, n. 5, p. 269–284, 2004.
29. MAURICE, T.; PHAN, V.-L.; URANI, A.; GUILLEMAIN, I. Differential involvement of the sigma₁ (σ_1) receptor in the anti-amnesic effect of neuroactive steroids, as demonstrated using an *in vivo* antisense strategy in the mouse. **British Journal of Pharmacology**, v. 134, n. 8, p. 1731–1741, 2001.

30. ROMERO, L.; ZAMANILLO, D.; NADAL, X.; SÁNCHEZ-ARROYOS, R.; RIVERA-ARCONADA, I.; DORDAL, A.; MONTERO, A.; MURO, A.; BURA, A.; SEGALÉS, C.; LALOYA, M.; HERNÁNDEZ, E.; PORTILLO-SALIDO, E.; ESCRICHE, M.; CODONY, X.; ENCINA, G.; BURGUEÑO, J.; MERLOS, M.; BAEYENS, J. M.; GIRALDO, J.; LÓPEZ-GARCÍA, J. A.; MALDONADO, R.; PLATA-SALAMÁN, C. R.; VELA, J. M. Pharmacological properties of S1RA, a new sigma-1 receptor antagonist that inhibits neuropathic pain and activity-induced spinal sensitization. **British Journal of Pharmacology**, v. 166, n. 8, p. 2289–2306, 2012.

31. 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 antagonist in three randomized phase I studies. **British Journal of Clinical Pharmacology**, v. 75, n. 1, p. 103–117, 2013.

32. MAVROMOUSTAKOS, T.; DURDAGI, S.; KOUKOULITSA, C.; SIMCIC, M.; PAPADOPOULOS, M. G.; HODOSCEK, M.; GRDADOLNIK, S. G. Strategies in the rational drug design. **Current Medicinal Chemistry**, v. 18, n. 17, p. 2517–2530, 2011.

33. LIAO, C.; SITZMANN, M.; PUGLIESE, A.; NICKLAUS, M. C. Software and resources for computational medicinal chemistry. **Future Medicinal Chemistry**, v. 3, n. 8, p. 1057–1085, 2011.

34. KUBINYI, H. The quantitative analysis of structure-activity relationships. In: WOLFF, M. E. (Ed.). **Burger's medicinal chemistry and drug discovery**. New York: John Wiley, 1995. v. 1, p. 497–571.

35. 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.

36. KUBINYI, H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. **Quantitative Structure-Activity Relationships**, v. 13, n. 4, p. 393–401, 1994.

37. LEARDI, R.; GONZÁLEZ, A. L. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. **Chemometrics and Intelligent Laboratory Systems**, v. 41, n. 2, p. 195–207, 1998.

38. WEBER, K. C.; DA SILVA, A. B. F. A chemometric study of the 5-HT(1A) receptor affinities presented by arylpiperazine compounds. **European Journal of Medicinal Chemistry**, v. 43, n. 2, p. 364–372, 2008.

39. DE MOLFETTA, F. A.; ANGELOTTI, W. F. D.; ROMERO, R. A. F.; MONTANARI, C. A.; DA SILVA, A. B. F. A neural networks study of quinone compounds with trypanocidal activity. **Journal of Molecular Modeling**, v. 14, n. 10, p. 975–985, 2008.
40. HONÓRIO, K. M.; DE LIMA, E. F.; QUILLES, M. G.; ROMERO, R. A. F.; MOLFETTA, F. A.; DA SILVA, A. B. F. Artificial neural networks and the study of the psychoactivity of cannabinoid compounds. **Chemical Biology & Drug Design**, v. 75, n. 6, p. 632–640, 2010.
41. LOZANO, N. B. H.; OLIVEIRA, R. F.; WEBER, K. C.; HONORIO, K. M.; GUIDO, R. V.; Andricopulo, A. D.; da Silva, A. B. F. Identification of electronic and structural descriptors of adenosine analogues related to inhibition of leishmanial glyceraldehyde-3-phosphate dehydrogenase. **Molecules**, v. 18, n. 5, p. 5032–5050, 2013.
42. TALEVI, A.; GOODARZI, M.; ORTIZ, E. V.; DUCHOWICZ, P. R.; BELLERA, C. L.; PESCE, G.; CASTRO, E. A.; BRUNO-BLANCH, L. E. Prediction of drug intestinal absorption by new linear and non-linear QSPR. **European Journal of Medicinal Chemistry**, v. 46, n. 1, p. 218–228, 2011.
43. GARKANI-NEJAD, Z.; AHMADI-ROUDI, B. Modeling the antileishmanial activity screening of 5-nitro-2-heterocyclic benzylidene hydrazides using different chemometrics methods. **European Journal of Medicinal Chemistry**, v. 45, n. 2, p. 719–726, 2010.
44. HASANI, M.; EMAMI, F. Evaluation of feed-forward back propagation and radial basis function neural networks in simultaneous kinetic spectrophotometric determination of nitroaniline isomers. **Talanta**, v. 75, n. 1, p. 116–126, 2008.
45. BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: a practical guide**. New York: Wiley, 1998. 348 p.
46. WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.
47. BRERETON, R. G. **Chemometrics: data analysis for the laboratory and chemical plant**. Chichester: John Wiley, 2003. 489 p.
48. KIRALJ, R.; FERREIRA, M. M. C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. **Journal of the Brazilian Chemical Society**, v. 20, n. 4, p. 770–787, 2009.

49. NOVIC, M.; VRACKO, M. Artificial neural networks in molecular structures-property studies. In: LEARDI, R. (Ed.). **Nature-inspired methods in chemometrics: genetic algorithms and artificial neural networks**. Amsterdam: Elsevier, 2003. p. 231–256.
50. HAYKIN, S. **Neural networks: a comprehensive foundation**. 2. ed. Upper Saddle River: Prentice Hall, 1999. 842 p.
51. ZUPAN, J. Basics of artificial neural networks. In: LEARDI, R. (Ed.). **Nature-inspired methods in chemometrics: genetic algorithms and artificial neural networks**. Amsterdam: Elsevier, 2003. p. 199–229.
52. 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, J. A. J.; PERALTA, J. E.; OGILARIO, 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, J. M.; 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 09**. Revisão A. 01. Wallingford: Gaussian, 2009. Programa de computador.
53. TETKO, I. V.; GASTEIGER, J.; TODESCHINI, R.; MAURI, A.; LIVINGSTONE, D.; ERTL, P.; PALYULIN, V. A.; RADCHENKO, E. V.; ZEFIROV, N. S.; MAKARENKO, A. S.; TANCHUK, V. Y.; PROKOPENKO, V. V. Virtual computational chemistry laboratory -- design and description. **Journal of Computer-Aided Molecular Design**, v. 19, n. 6, p. 453–463, 2005.
54. 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.
55. PIROUETTE. Versão 3.10. [S.I.]: Infometrix, 2002. Programa de computador.
56. 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.

57. MATLAB. Versão 7.12. [S.I.]: MathWorks, 2011. Programa de computador.
58. GOLBRAIKH, A.; TROPSHA, A. Beware of q^2 ! **Journal of Molecular Graphics and Modelling**, v. 20, n. 4, p. 269–276, 2002.
59. 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.
60. WOLD, S.; ERIKSSON, L.; CLEMENTI, S. Statistical validation of QSAR results. In: WATERBEEMD, H. VAN DE (Ed.). **Chemometric methods in molecular design**. New York: VCH, 1995. p. 309–338.
61. ERIKSSON, L.; JAWORSKA, J.; WORTH, A. P.; CRONIN, M. T. D.; MCDOWELL, R. M.; GRAMATICA, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. **Environmental Health Perspectives**, v. 111, n. 10, p. 1361–1375, 2003.
62. ZHANG, S.; GOLBRAIKH, A.; OLOFF, S.; KOHN, H.; TROPSHA, A. A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications and virtual screening of chemical databases using validated ALL-QSAR models. **Journal of Chemical Information and Modeling**, v. 46, n. 5, p. 1984–1995, 2006.
63. GLENNON, R. A.; ABLORDEPPEY, S. Y.; ISMAIEL, A. B.; EL-ASHMAWY, M. B.; FISCHER, J. B.; HOWIE, K. B. Structural Features Important for σ_1 Receptor Binding. **Journal of Medicinal Chemistry**, v. 37, n. 8, p. 1214–1219, 1994.
64. GÁLVEZ, J.; GARCIA, R.; SALABERT, M. T.; SOLER, R. Charge indexes. New topological descriptors. **Journal of Chemical Information and Computer Sciences**, v. 34, n. 3, p. 520–525, 1994.
65. GÁLVEZ, J.; GARCÍA-DOMENECH, R.; DE JULIÁN-ORTIZ, J. V.; SOLER, R. Topological approach to drug design. **Journal of Chemical Information and Computer Sciences**, v. 35, n. 2, p. 272–284, 1995.
66. FORESMAN, J. B.; FRISCH, AE. **Exploring chemistry with electronic structure methods**. 2. ed. Pittsburgh: Gaussian, 1996. 302 p.

67. EROGLU, E.; TÜRKMEN, H.; GÜLER, S.; PALAZ, S.; OLTULU, O. A DFT-based QSARs study of acetazolamide/sulfanilamide derivatives with carbonic anhydrase (CA-II) isozyme inhibitory activity. **International Journal of Molecular Sciences**, v. 8, n. 2, p. 145–155, 2007.
68. CLARE, B. W.; SUPURAN, C. T. Semi-empirical atomic charges and dipole moments in hypervalent sulfonamide molecules: descriptors in QSAR studies. **Journal of Molecular Structure: THEOCHEM**, v. 428, n. 1-3, p. 109–121, 1998.
69. RUECKER, G.; RUECKER, C. Counts of all walks as atomic and molecular descriptors. **Journal of Chemical Information and Computer Sciences**, v. 33, n. 5, p. 683–695, 1993.
70. DIUDEA, M. V.; TOPAN, M.; GRAOVAC, A. Molecular topology. 17. Layer matrices of walk degrees. **Journal of Chemical Information and Computer Sciences**, v. 34, n. 5, p. 1072–1078, 1994.
71. SOLTZBERG, L. J.; WILKINS, C. L. Molecular transforms : a potential tool for structure-activity studies. **Journal of the American Chemical Society**, v. 99, n. 2, p. 439–443, 1977.
72. SCHUUR, J. H.; SELZER, P.; GASTEIGER, J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. **Journal of Chemical Information and Computer Sciences**, v. 36, n. 2, p. 334–344, 1996.
73. GASTEIGER, J.; SADOWSKI, J.; SCHUUR, J.; SELZER, P.; STEINHAEUER, L.; STEINHAEUER, V. Chemical information in 3D space. **Journal of Chemical Information and Computer Sciences**, v. 36, n. 5, p. 1030–1037, 1996.
74. CONSONNI, V.; TODESCHINI, R.; PAVAN, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 3, p. 682–692, 2002.
75. CONSONNI, V.; TODESCHINI, R.; PAVAN, M.; GRAMATICA, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 3, p. 693–705, 2002.

76. MORIGUCHI, I.; HIRONO, S.; LIU, Q.; NAKAGOME, I.; MATSUSHITA, Y. Simple method of calculating octanol/water partition coefficient. **Chemical and Pharmaceutical Bulletin**, v. 40, n. 1, p. 127–130, 1992.

77. KUJAWSKI, J.; POPIELARSKA, H.; DRABINSKA, A. M. B.; BERNARD, M. K. The log P parameter as a molecular descriptor in the computer-aided drug design – an overview. **Computational Methods in Science and Technology**, v. 18, n. 2, p. 81–88, 2012.

APÊNDICE

APÊNDICE A. Parâmetros estatísticos dos modelos MLP-ANN obtidos.

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>tansig-tansig</i>	1	0,0079	0,8095	0,0691	6	0,0079	0,8107	0,0692	11	0,0082	0,8037	0,0689	16	0,0083	0,8014	0,0668
	2	0,0080	0,8071	0,0701	7	0,0089	0,7860	0,0699	12	0,0070	0,8309	0,0626	17	0,0080	0,8084	0,0650
	3	0,0082	0,8022	0,0692	8	0,0080	0,8069	0,0643	13	0,0078	0,8129	0,0678	18	0,0085	0,7957	0,0698
	4	0,0075	0,8190	0,0656	9	0,0087	0,7916	0,0708	14	0,0087	0,7916	0,0713	19	0,0083	0,8005	0,0673
	5	0,0074	0,8227	0,0674	10	0,0085	0,7945	0,0681	15	0,0080	0,8085	0,0708	20	0,0088	0,7891	0,0736
<i>logsig-logsig</i>	1	0,0158	0,6207	0,0866	6	0,0153	0,6321	0,0815	11	0,0159	0,6175	0,0869	16	0,0166	0,6009	0,0845
	2	0,0153	0,6331	0,0814	7	0,0164	0,6044	0,0864	12	0,0162	0,6097	0,0870	17	0,0163	0,6078	0,0868
	3	0,0154	0,6304	0,0849	8	0,0166	0,6002	0,0900	13	0,0162	0,6100	0,0854	18	0,0166	0,6018	0,0841
	4	0,0156	0,6253	0,0833	9	0,0160	0,6141	0,0864	14	0,0150	0,6387	0,0769	19	0,0163	0,6089	0,0834
	5	0,0163	0,6084	0,0867	10	0,0150	0,6400	0,0739	15	0,0160	0,6143	0,0873	20	0,0161	0,6125	0,0846
<i>purelin-purelin</i>	1	0,0109	0,7377	0,0818	6	0,0111	0,7336	0,0822	11	0,0119	0,7148	0,0847	16	0,0105	0,7465	0,0813
	2	0,0110	0,7366	0,0820	7	0,0111	0,7336	0,0822	12	0,0111	0,7336	0,0822	17	0,0102	0,7542	0,0756
	3	0,0109	0,7369	0,0818	8	0,0137	0,6700	0,0910	13	0,0109	0,7376	0,0819	18	0,0101	0,7582	0,0767
	4	0,0111	0,7338	0,0822	9	0,0111	0,7336	0,0822	14	0,0109	0,7388	0,0810	19	0,0130	0,6878	0,0862
	5	0,0110	0,7360	0,0823	10	0,0111	0,7336	0,0822	15	0,0112	0,7307	0,0821	20	0,0132	0,6828	0,0852
<i>tansig-logsig</i>	1	0,0161	0,6137	0,0877	6	0,0142	0,6591	0,0751	11	0,0140	0,6630	0,0724	16	0,0163	0,6072	0,0835
	2	0,0152	0,6342	0,0815	7	0,0158	0,6206	0,0810	12	0,0162	0,6106	0,0775	17	0,0142	0,6591	0,0669
	3	0,0158	0,6197	0,0845	8	0,0146	0,6479	0,0738	13	0,0138	0,6692	0,0651	18	0,0177	0,5753	0,0914
	4	0,0157	0,6214	0,0816	9	0,0146	0,6498	0,0701	14	0,0149	0,6409	0,0778	19	0,0155	0,6259	0,0812
	5	0,0156	0,6254	0,0807	10	0,0165	0,6031	0,0831	15	0,0152	0,6347	0,0683	20	0,0157	0,6222	0,0716

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>tansig-purelin</i>	1	0,0079	0,8106	0,0689	6	0,0076	0,8167	0,0648	11	0,0079	0,8104	0,0658	16	0,0081	0,8055	0,0710
	2	0,0082	0,8033	0,0701	7	0,0077	0,8143	0,0665	12	0,0087	0,7910	0,0735	17	0,0074	0,8212	0,0627
	3	0,0072	0,8280	0,0655	8	0,0081	0,8054	0,0697	13	0,0063	0,8493	0,0623	18	0,0086	0,7940	0,0717
	4	0,0065	0,8433	0,0610	9	0,0075	0,8200	0,0543	14	0,0081	0,8059	0,0667	19	0,0104	0,7506	0,0739
	5	0,0075	0,8197	0,0640	10	0,0084	0,7968	0,0697	15	0,0083	0,7993	0,0638	20	0,0085	0,7951	0,0678
<i>logsig-tansig</i>	1	0,0079	0,8097	0,0690	6	0,0076	0,8181	0,0656	11	0,0080	0,8080	0,0660	16	0,0080	0,8068	0,0664
	2	0,0081	0,8051	0,0701	7	0,0076	0,8173	0,0673	12	0,0079	0,8095	0,0667	17	0,0086	0,7934	0,0641
	3	0,0079	0,8065	0,0696	8	0,0078	0,8113	0,0684	13	0,0077	0,8142	0,0643	18	0,0082	0,8020	0,0743
	4	0,0080	0,8274	0,0676	9	0,0073	0,8239	0,0636	14	0,0082	0,8039	0,0712	19	0,0079	0,8106	0,0627
	5	0,0076	0,8179	0,0655	10	0,0079	0,8111	0,0653	15	0,0088	0,7874	0,0647	20	0,0074	0,8226	0,0650
<i>logsig-purelin</i>	1	0,0079	0,8103	0,0689	6	0,0082	0,8038	0,0688	11	0,0075	0,8188	0,0676	16	0,0089	0,7862	0,0692
	2	0,0078	0,8117	0,0696	7	0,0073	0,8235	0,0620	12	0,0078	0,8115	0,0638	17	0,0077	0,8153	0,0688
	3	0,0079	0,8097	0,0691	8	0,0079	0,8098	0,0644	13	0,0075	0,8204	0,0667	18	0,0086	0,7928	0,0668
	4	0,0078	0,8125	0,0679	9	0,0075	0,8192	0,0672	14	0,0089	0,7864	0,0729	19	0,0079	0,8100	0,0653
	5	0,0075	0,8197	0,0663	10	0,0086	0,7938	0,0693	15	0,0079	0,8108	0,0643	20	0,0069	0,8331	0,0616
<i>purelin-tansig</i>	1	0,0086	0,7926	0,0719	6	0,0087	0,7898	0,0736	11	0,0087	0,7911	0,0716	16	0,0086	0,7933	0,0728
	2	0,0086	0,7929	0,0722	7	0,0087	0,7915	0,0728	12	0,0085	0,7960	0,0720	17	0,0088	0,7878	0,0741
	3	0,0087	0,7896	0,0735	8	0,0087	0,7900	0,0727	13	0,0088	0,7887	0,0738	18	0,0087	0,7907	0,0725
	4	0,0087	0,7897	0,0729	9	0,0088	0,7887	0,0738	14	0,0087	0,7912	0,0727	19	0,0088	0,7883	0,0739
	5	0,0087	0,7907	0,0733	10	0,0087	0,7909	0,0731	15	0,0088	0,7890	0,0721	20	0,0088	0,7877	0,0741
<i>purelin-logsig</i>	1	0,0189	0,5449	0,1009	6	0,0190	0,5426	0,1007	11	0,0190	0,5422	0,1005	16	0,0190	0,5435	0,1012
	2	0,0189	0,5462	0,1008	7	0,0190	0,5431	0,1007	12	0,0191	0,5414	0,1012	17	0,0189	0,5445	0,1007
	3	0,0190	0,5419	0,1007	8	0,0190	0,5438	0,1007	13	0,0190	0,5419	0,1007	18	0,0188	0,5469	0,1006
	4	0,0190	0,5441	0,1007	9	0,0189	0,5453	0,1007	14	0,0190	0,5431	0,1005	19	0,0191	0,5406	0,1009
	5	0,0190	0,5432	0,1006	10	0,0188	0,5477	0,1009	15	0,0189	0,5459	0,1007	20	0,0190	0,5425	0,1006

*Número de neurônios na camada intermediária

**CAPÍTULO 2: Proposição de bicamadas lipídicas
contendo POPC, DOPE, POPS e colesterol a
partir de simulações por dinâmica molecular**

1 INTRODUÇÃO

1.1 Membranas celulares

Quando glicerofosfolípidos, esfingolípidos e esteróides são expostos à água, eles se agregam de forma que os grupos hidrofóbicos fiquem em contato entre si e os grupos hidrofílicos sejam expostos à água. Essa agregação de lípidos pode gerar três formas distintas: micela, bicamada ou lipossoma (Figura 1). As micelas são estruturas esféricas em que as regiões hidrofílicas estão voltadas para seu exterior enquanto que a parte hidrofóbica está no seu interior, onde não há presença de moléculas de água. Já nas bicamadas, há formação de duas monocamadas de lípidos cuja parte hidrofóbica de uma monocamada interage com a parte hidrofóbica da outra monocamada e a porção hidrofílica interage com a água na superfície da bicamada. Em certas situações, os lípidos da borda da bicamada que estão expostos à água, e por isso são relativamente instáveis, podem se agregar formando lipossomas. Esse tipo de agregado mantém moléculas de água em seu interior.^{1,2}

Figura 1 - Agregados de lípidos formados em água: micela, bicamada e lipossoma.

Fonte: Adaptação a partir de BITOUNIS, D. *et al.*, 2012, p. 2.

Dentre esses agregados, a bicamada lipídica, juntamente com proteínas de transporte e receptoras e enzimas de membrana, constitui a membrana celular. As bicamadas lipídicas são formadas por uma mistura complexa de glicerofosfolípidos, esfingolípidos e esteróides que são distribuídos de maneira assimétrica entre as duas monocamadas. As membranas celulares de mamíferos contêm diferentes espécies de lípidos das classes dos glicerofosfolípidos e esfingolípidos e, principalmente, o colesterol da classe dos esteróides.^{1,3,4}

A estrutura de membranas celulares pode ser explicada pelo modelo do mosaico fluido proposto por Singer e Nicolson em 1972. Esse modelo descreve a membrana como uma bicamada fluida composta por lipídeos, proteínas integrais de membrana e glicoproteínas (Figura 2).³ Uma vez que grande parte das interações entre os componentes são não covalentes, o modelo de membrana adquire fluidez já que as moléculas de lipídeo e as proteínas podem se mover livremente no plano da membrana. Embora tenha sido o modelo de maior sucesso para descrever membranas biológicas, o modelo do mosaico fluido vem sendo complementado para incluir novas descobertas como agregação e segregação de proteínas e lipídeos em domínios e interações com o meio extracelular e com o citoesqueleto.⁵

Figura 2 - Modelo do mosaico fluido proposto por Singer e Nicolson em 1972. As estruturas sólidas representam as proteínas integrais de membrana que estão distribuídas em uma bicamada lipídica fluida.

Fonte: NICOLSON, G. L., 2014, p. 1452.

Além de alterar a conformação e a dinâmica dos receptores que estão acoplados à membrana, os lipídeos podem desempenhar papéis fundamentais nos mecanismos de interação com ligantes. Estudos desenvolvidos por Pöyry *et al.* (2013), por exemplo, sugerem que moléculas de cardiolipina presentes na membrana estão associadas à estrutura do complexo do citocromo bc1, além de favorecerem o processo de transferência de prótons.⁶

Alguns ligantes podem ainda interagir com as membranas, provocando alterações em suas propriedades biofísicas. As propriedades farmacológicas de anestésicos, anti-inflamatórios não esteroidais, anti-psicóticos, dentre outras pequenas moléculas, podem estar associadas também às interações que eles estabelecem com a bicamada lipídica.⁷

Acredita-se que essa interação entre ligantes e seus receptores acoplados às membranas possa acontecer de duas formas distintas: eles podem interagir diretamente com o

alvo (difusão 3D) ou podem interagir com a membrana por meio de um processo de adsorção reversível onde se difundem até atingir o alvo (difusão 3D-2D).^{8,9}

Portanto, estudos detalhados da composição da membrana tornam-se essenciais para a compreensão do mecanismo de ação de determinado composto. Os lipídeos que compõem a bicamada podem afetar a interação com as pequenas moléculas, provocando assim um aumento ou diminuição da sua propriedade farmacológica.

Recentemente, Postila e colaboradores (2016) estudaram a interação entre treze neurotransmissores e três ambientes membranares através de simulações por DM. Os resultados indicam que alguns dos neurotransmissores estudados estabelecem interações importantes com a bicamada lipídica, o que reforça a proposta de que eles se ligam primeiro à membrana por onde se deslocam até atingir o sítio de ligação de determinado receptor sináptico. Eles ainda sugerem que a composição da bicamada lipídica, em especial os lipídeos carregados, pode afetar diretamente a interação com os neurotransmissores.¹⁰

O estudo das interações entre membrana e neurotransmissores é fundamental para a compreensão de doenças como Alzheimer, Parkinson e esquizofrenia. Alguns estudos já têm sugerido que tais doenças estão associadas a uma alteração da composição de lipídeos das membranas sinápticas.¹¹ Portanto, a composição das membranas sinápticas e o mecanismo de interação das mesmas com receptores ou mesmo com neurotransmissores tornam-se fatores importantes na busca de tratamentos para essas doenças.

De acordo com Cotman *et al.* (1969), as membranas sinápticas de mamíferos têm em sua composição fosfatidilcolina (34%), fosfatidiletanolamina (28%), fosfatidilserina e fosfatidilinositol (12%), esfingomielina (3%), colesterol (18%) e lisofosfatidilcolina (1%).¹² Entretanto, a composição lipídica das membranas sinápticas se altera com o avanço da idade ou pode ser induzida por estados degenerativos de doenças como o Alzheimer.^{11,13,14} Estudos com camundongos durante quatro diferentes estágios de vida têm indicado que no estágio jovem a distribuição do colesterol é de 15%/85% entre as superfícies externa e interna da membrana sináptica, respectivamente. Com o avanço da idade, essa distribuição pode chegar a 50%/50%.¹⁵

O colesterol presente nas membranas celulares desempenha funções associadas à sinalização celular, neurotransmissão e organização das proteínas na membrana.¹⁶⁻¹⁸ Além disso, tem sido observado que o colesterol é capaz de afetar receptores, transportadores e canais iônicos de forma direta, por meio de ligação com a proteína, indiretamente quando o

mesmo altera as propriedades da membrana ou até mesmo uma combinação dos dois modos de ação.¹⁹⁻²¹

1.1.1 Influência do colesterol nas propriedades físico-químicas das membranas

Dentre as propriedades das bicamadas lipídicas, a sua fluidez é uma das mais importantes a ser analisada. Abaixo da temperatura de transição de fase, a membrana encontra-se na fase gel (L_β) onde há menor mobilidade. À medida que a temperatura é aumentada, a membrana apresenta maior mobilidade. Após atingir a temperatura de transição de fase, a membrana atinge a fase líquido-cristalina (L_α). A fase L_α tem sido amplamente estudada, uma vez que nela acontecem muitos processos biológicos importantes.²²

As temperaturas de transição de fase são diferentes para cada fosfolípídeo. Para as moléculas de POPC (1-palmitoil-2-oleoil-sn-glicero-3-fosfocolina), por exemplo, a temperatura de transição de fase é de 271 K, enquanto que para DOPE (1,2-dioleoil-sn-glicero-3-fosfoetanolamina) e POPS (1-palmitoil-2-oleoil-sn-glicero-3-fosfoserina) os valores das temperaturas de transição são de 257 e 287 K, respectivamente.²³

As propriedades estruturais das bicamadas lipídicas podem ser alteradas significativamente pela presença de moléculas de colesterol. A estrutura planar e rígida da molécula de colesterol e sua pequena cabeça polar favorecem o contato com as cadeias acil dos fosfolípídeos, o que permite que haja alterações estruturais da bicamada lipídica. Diversos estudos têm mostrado que moléculas de colesterol de uma bicamada lipídica na fase L_α induzem a um ordenamento das cadeias alifáticas.²⁴⁻²⁶ Além disso, o colesterol induz um aumento da espessura da bicamada (EB) e uma redução da permeabilidade da membrana.²⁷⁻²⁹ As interações lipídeo-lipídeo e lipídeo-proteína também podem ser afetadas pela presença de moléculas de colesterol.³⁰

Outros estudos têm sugerido ainda que o colesterol é capaz de induzir a formação de domínios (*lipid rafts*) em bicamadas lipídicas contendo fosfatidilcolinas ou esfingolípídeos compostos por cadeias alifáticas saturadas. Tais domínios são fluidos, mas mais ordenados do que o sistema a sua volta.^{31,32} Eles desempenham papéis importantes no processo de transdução de sinal, polarização da célula e na distribuição de proteínas e lipídeos na membrana. Além disso, existem várias doenças associadas a esses domínios, tais como: Alzheimer, Parkinson, diabetes, dentre muitas outras.³³

Em bicamadas complexas que contêm moléculas de colesterol observam-se ainda as fases líquido-ordenada (L_o) e líquido-desordenada (L_d), sendo que a fase L_o corresponde a um estado de fluidez intermediária entre as fases gel e fases fluidas. Os diagramas de fases propostos para misturas binárias de lipídeo/colesterol mostram que as misturas possuem no mínimo uma fase L_o , e que há regiões onde as fases L_o e L_d coexistem.^{34,35}

1.1.2 Bicamadas lipídicas e as simulações por DM

A primeira simulação de um peptídeo inserido em uma bicamada lipídica foi realizada em 1994 por Woolf e Roux. O canal de gramicidina foi inserido em uma bicamada de DMPC (1,2-dimiristoil-sn-glicero-3-fosfocolina) com o objetivo de estudar as interações lipídeo-proteína. As simulações mostraram as cadeias laterais de triptofano interagindo com o grupo éster carbonila das moléculas de DMPC e com a água por meio de ligações hidrogênio. Tais interações são fundamentais para a estabilidade de proteínas de membrana.³⁶

Já em 1995, foi feita a primeira simulação de uma proteína integral de membrana. Edholm e colaboradores estudaram a bacteriorrodopsina inserida em uma bicamada lipídica composta por difitanil fosfatidil-glicerol-fosfato. Os resultados evidenciaram uma flutuação assimétrica da proteína em relação ao plano médio da bicamada.³⁷

Liguori e colaboradores (2013), na tentativa de estudar interações entre membranas sinápticas e o peptídeo β -amiloide, simularam uma bicamada lipídica utilizando moléculas de POPC, DOPE, POPS e colesterol. Nesse estudo, os autores avaliaram diferentes concentrações e distribuição do colesterol e seu efeito sobre o peptídeo β -amiloide, através de simulações por DM utilizando o modelo de *coarse-grained* MARTINI.³⁸ Recentemente, Cao e colaboradores (2015) desenvolveram simulações do receptor de adenosina em bicamadas lipídicas utilizando o modelo *all-atom* Slipids. Entretanto, eles utilizaram apenas POPC, POPE e colesterol para simular o ambiente de uma membrana sináptica.³⁹

Nos estudos citados anteriormente e em diversos outros disponíveis na literatura, as simulações por DM forneceram informações relevantes para a compreensão dos sistemas investigados. O comportamento das bicamadas lipídicas durante as simulações por DM é avaliado observando-se diferentes parâmetros estruturais. Com base nessas análises, é possível identificar o efeito provocado por diferentes concentrações de lipídeos, por diferentes temperaturas, pela presença de pequenos ligantes, peptídeos ou mesmos de proteínas.^{38,40,41}

Dentre os parâmetros estruturais que têm sido avaliados para esse propósito, destacam-se a área por lipídeo (APL), EB, difusão lateral (DL), parâmetro de ordem (PO), dentre outros.

A APL é sensível a alterações da bicamada lipídica, tais como o comprimento das cadeias alifáticas, a temperatura ou mesmo a composição lipídica.^{27,40,42,43} Analisando-se a APL de uma bicamada lipídica simulada por DM, é possível determinar se o sistema está bem equilibrado, uma vez que a APL tende a ficar constante. Para simulações de bicamadas lipídicas que não tenham o valor experimental de APL, como nos casos de muitas bicamadas mistas, a análise de APL é usada apenas para avaliar a estabilidade do sistema, sem considerar seu valor exato.

Outra importante propriedade das bicamadas lipídicas e que tem sido determinada de diferentes maneiras é a EB. O espalhamento de raios-X e o espalhamento de neutrôns são exemplos de técnicas experimentais muito utilizadas para a determinação da EB, mas que fornecem valores distintos. Os valores de EB determinados pela técnica de espalhamento de raios-X correspondem à distância entre os grupos fosfatos dos lipídeos, obtida a partir da distância entre picos em um gráfico do perfil de densidade eletrônica. Já a técnica de espalhamento de neutrôns determina os valores de EB a partir do alto contraste entre lipídeos protonados e água deuterada.⁴³

É possível ainda obter a DL de uma bicamada lipídica a partir do seu deslocamento quadrático médio determinado por diferentes métodos como espalhamento de nêutrons e fluorescência. Esses métodos produzem resultados diferentes para os valores do coeficiente de difusão, sendo que para tempos mais longos esses valores podem chegar a duas ordens de magnitudes menores do que aqueles obtidos para tempos curtos. Simulações por DM de bicamadas de DPPC (1,2-dipalmitoil-sn-glicero-3-fosfocolina) mostram que os valores do coeficiente de difusão para 10 ns de simulação estão em acordo com valores experimentais determinados pela técnica de espalhamento de nêutrons.⁴⁴ Entretanto, ao simular 100 ns do mesmo sistema, o coeficiente de difusão é menor do que para a simulação mais curta e se assemelha mais aos resultados experimentais determinados pela técnica de fluorescência.⁴⁵

Já para avaliar o grau de ordenamento de uma bicamada lipídica, determina-se o PO a partir da equação 1. Um estado ordenado tem valor absoluto de PO igual a 1 e quanto mais próximo de zero for o valor de PO, menor será o alinhamento das cadeias alifáticas.

$$S_{CD} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle \quad (1)$$

sendo que θ é o ângulo entre o vetor da ligação C-D (experimento) ou C-H (simulação) e a normal à bicamada.⁴⁶

A determinação dessas propriedades estruturais e de várias outras obtidas a partir das simulações por DM de bicamadas lipídicas permite avaliar os sistemas simulados e investigar se as propriedades calculadas correspondem aos valores experimentais disponíveis para o sistema de interesse.

1.2 Mecânica Molecular

Para realizar os estudos de sistemas com um grande número de átomos, tais como membranas biológicas, faz-se necessário utilizar os métodos baseados em mecânica molecular (MM). Nesses métodos utiliza-se um campo de forças que descreve a energia potencial de um sistema a partir de equações referentes às interações intra e intermoleculares entre as partículas. Entretanto, o movimento eletrônico é desconsiderado o que impede que fenômenos como formação ou ruptura de ligações químicas, por exemplo, sejam observados.⁴⁷

A descrição de oscilações no comprimento de ligações químicas em um campo de forças pode ser dada por um potencial aproximado da lei de Hooke (eq. 2).

$$v_i = \frac{1}{2} \sum_{n=1}^{N_b} k_{bn} (r_n - r_{i,0})^2 \quad (2)$$

em que N_b refere-se ao número total de pares de átomos ligados, k_{bn} representa uma constante de força, r_n é a distância entre os átomos e $r_{i,0}$ é a distância de referência.

Já o ângulo formado entre duas ligações consecutivas pode ser descrito pelo seguinte potencial harmônico (eq. 3):

$$v_\theta = \frac{1}{2} \sum_{n=1}^{N_\theta} k_{\theta n} (\theta_n - \theta_{i,0})^2 \quad (3)$$

sendo N_θ o número total de sequências de três átomos, $k_{\theta n}$ a constante de força, θ_n o ângulo formado entre os átomos e $\theta_{i,0}$ o ângulo de equilíbrio.

Para a representação dos potenciais torcionais próprios e impróprios pode-se utilizar séries de cossenos, tal como descrito nas equações 4 e 5, respectivamente.

$$v(\emptyset) = \sum_{n=0}^N \frac{V_{\emptyset n}}{2} [1 + \cos(m\emptyset_n - y_n)] \quad (4)$$

$$v(w) = k (1 - \cos 2w) \quad (5)$$

em que $\emptyset$ refere-se ao ângulo de torção, $V_{\emptyset n}$ representa a constante da barreira de torção, m é o número mínimo de pontos na função quando há uma rotação de 360° na ligação, y_n é o fator de fase associado à determinação do ponto de mínimo ou de máximo, w é o ângulo torcional impróprio.

Para representar as interações que acontecem entre os átomos não ligados, utilizam-se potenciais capazes de descrever, por exemplo, interações de van der Waals e eletrostáticas. Para descrever as interações de van der Waals, é comum utilizar uma equação que descreve tanto termos repulsivos quanto atrativos. Um exemplo está descrito na equação 6.

$$v_{vdw} = \sum_{i < j}^{N^*} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (6)$$

sendo que as interações são calculadas para os átomos i e j de um conjunto total de N^* átomos, onde aqueles que são tratados pelos potenciais harmônicos são excluídos e os terceiros vizinhos são tratados com parâmetros reduzidos no potencial. A_{ij} e B_{ij} incluem parâmetros que representam energia e diâmetro de Lennard-Jones ($4\epsilon\sigma^{12}$ e $4\epsilon\sigma^6$, respectivamente) e r_{ij} corresponde à distância entre os pares de átomos.

Para descrever as interações eletrostáticas, pode-se utilizar um potencial da forma da equação 7:

$$v_{el} = \sum_{i < j}^{N^*} \frac{q_i q_j}{4\pi\epsilon\epsilon_0 r_{ij}} \quad (7)$$

sendo que q_i e q_j representam as cargas parciais dos átomos e r_{ij} a distância entre eles. ε e ε_0 são a constante dielétrica do meio e a constante de permissividade no vácuo, respectivamente.

Uma vez descritos os potenciais mais comuns utilizados em um campo de forças, os mesmos são combinados em uma equação que vai determinar a energia potencial total ($v(r^N)$) em função das posições ($\vec{r}$) de um sistema de partículas (eq. 8)

$$\begin{aligned}
 v(r^N) = & \sum_{\text{ligações}} \frac{k_i}{2} (r_i - r_{i,0})^2 \\
 & + \sum_{\text{ângulos}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \\
 & + \sum_{\text{torções}} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] + \sum_{i=1}^N \sum_{j=1}^N \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad (8)
 \end{aligned}$$

Na equação 8 o primeiro termo representa o somatório das interações entre os átomos que estão ligados diretamente, enquanto que o segundo termo está associado aos ângulos que se formam entre duas ligações consecutivas. O terceiro termo descreve o potencial torsional gerado a partir do giro das ligações químicas. As interações entre os átomos não ligados que estão em moléculas diferentes ou na mesma molécula, mas distantes por no mínimo três ligações, são descritas no quarto e quinto termos, onde têm-se o potencial de Lennard-Jones que descreve as interações de van der Waals e o potencial de Coulomb que representa as interações eletrostáticas, respectivamente.⁴⁷

1.2.1 Minimização de energia

Ao se estudar sistemas biológicos que possuam um alto número de graus de liberdade, torna-se necessário realizar previamente uma otimização de geometria que encontre um conjunto de coordenadas que minimizem a energia potencial do sistema. A equação 9 representa o processo iterativo empregado na minimização de energia.⁴⁷

$$r_{i,n+1} = r_{i,n} + \Delta r_{i,n} \quad (9)$$

sendo n o número de iterações, $r_{i,n+1}$ as coordenadas do átomo i no passo atual, $r_{i,n}$ as coordenadas do átomo i no passo anterior e $\Delta r_{i,n}$ o incremento nas coordenadas do átomo i .

Diferentes métodos de minimização de energia são utilizados para determinar o módulo e a direção do passo $\Delta r_{i,n}$. Para as configurações que estejam muito distantes do mínimo de energia, é comum utilizar o método de primeira derivada *steepest-descent*. Entretanto, como esse método tem uma convergência lenta próximo ao mínimo de energia, nessa região utiliza-se métodos de segunda derivada como o de gradientes conjugados que é capaz de considerar o caminho que já foi percorrido até se atingir o mínimo de energia. Há ainda métodos de segunda ordem que utilizam primeira e segunda derivadas para atingir o mínimo de energia, tal como o método Newton-Raphson.⁴⁷

1.2.2 Dinâmica molecular

A DM tem sido muito utilizada atualmente a fim de se estudar a evolução temporal de um sistema. A partir de um potencial que descreva as interações entre as partículas e de suas equações de movimento, é possível simular o comportamento dessas partículas durante um tempo previamente determinado. Uma vez simulado esse sistema, muitas informações de interesse poderão ser obtidas a partir de sua trajetória, tais como: energia, temperatura, desvio de raiz média quadrática, número de ligações hidrogênio formadas, raio de giro, dentre muitas outras.

Inicialmente, constrói-se o sistema de interesse que será confinado em uma caixa de simulação com formato definido pelo usuário (ortorrômbico, cúbico, triclínico, dentre outros). A essa caixa são impostas condições periódicas de contorno para que o movimento das partículas não seja limitado à sua fronteira. São criadas réplicas idênticas do sistema original que são inseridas ao redor do mesmo produzindo assim um sistema em que o número de moléculas e o volume tendem ao infinito, mas a densidade numérica permanece constante. Com isso, todos os movimentos que acontecem no sistema original acontecem também em cada réplica.

Antes de se realizar uma simulação por DM é preciso ainda definir as velocidades iniciais dos átomos do sistema que, posteriormente, possibilitarão a resolução das equações de movimento. Independente do método escolhido para a geração das velocidades iniciais, a soma das velocidades sobre todas as partículas deverá ser um vetor nulo para que a caixa de simulação não se desloque como um todo. Dentre as diferentes formas disponíveis para se determinar as velocidades iniciais, pode-se utilizar a distribuição de Maxwell-Boltzmann (eq.

10) na qual se determina a energia cinética do sistema a partir de um determinado valor de temperatura.

$$\frac{3}{2} N k_B T = \frac{1}{2} \sum_{i=1}^N m_i v_i^2 \quad (10)$$

em que N representa o número de partículas, k_B é a constante de Boltzmann, T representa a temperatura e m_i e v_i são, respectivamente, a massa e a velocidade da partícula i .

Uma vez definidas as posições e as velocidades iniciais das partículas de um sistema, a resolução de equações de movimento permitirá determinar as posições e as velocidades em uma etapa posterior. O sistema isolado será então representado pela Hamiltoniana da equação 11:

$$H_N(r^N, p^N) = \frac{1}{2m} \sum_{i=1}^N |p_i|^2 + V_N(r^N) \quad (11)$$

sendo a soma das energias potencial e cinética do sistema uma constante. As coordenadas das partículas em instantes posteriores poderão então ser calculadas a partir das posições (r) e do momento (p) iniciais, resolvendo-se as equações de movimento de Newton (eq. 12) que consistem em equações diferenciais de segunda ordem acopladas.⁴⁸

$$0 = m\ddot{r}_i + \nabla_i V_N(r^N) \quad (12)$$

Existem vários programas disponíveis para realizar as simulações por DM. Dentre eles destacam-se AMBER⁴⁹, CHARMM⁵⁰, GROMACS⁵¹ e NAMD⁵². Para as simulações de bicamadas lipídicas, diferentes campo de forças são utilizados, desde os modelos *coarse-grained* como Martini⁵³ até modelos *all-atom* descritos por campo de forças como CHARMM^{54,55}, AMBER FF⁵⁶ e GLYCAM06⁵⁷.

Simulações por DM têm sido muito aplicadas no estudo de bicamadas lipídicas com o propósito de investigar a influência de moléculas como o colesterol ou outros lipídeos nas propriedades estruturais das bicamadas ou mesmo para avaliar o comportamento de pequenos peptídeos ou proteínas inseridos na bicamada lipídica.^{38,40,58}

Entretanto, ainda torna-se necessário desenvolver simulações em que as composições das bicamadas lipídicas sejam cada vez mais próximas das composições encontradas nos sistemas biológicos. Uma vez que a composição de uma bicamada é capaz de afetar diretamente as interações que acontecem entre os lipídeos, entre lipídeos e proteínas ou peptídeos ou mesmo entre a bicamada e algumas moléculas pequenas, é fundamental a busca por sistemas cada vez mais realísticos.

As simulações por DM podem contribuir significativamente para o avanço das pesquisas científicas dedicadas aos sistemas membranares. Nesse propósito, o presente trabalho busca propor modelos de bicamadas lipídicas que se assemelham à composição experimental determinada para as membranas sinápticas. Além disso, a utilização do campo de forças CHARMM36 permitirá descrever todos os átomos presentes no sistema, o que contribui para estudos futuros de interações que podem acontecer entre bicamadas lipídicas e receptores ou moléculas como os neurotransmissores. Uma vez que serão propostas estruturas de bicamadas lipídicas com diferentes concentrações de colesterol, será possível ainda disponibilizar modelos de membranas que permitirão avaliar a influência do colesterol na interação com pequenas moléculas como os neurotransmissores, o que contribui para o estudo de doenças como Alzheimer e Parkinson.

2 OBJETIVOS

Os estudos desenvolvidos têm por objetivo principal propor estruturas de bicamadas lipídicas que se assemelhem à composição de membranas sinápticas com diferentes concentrações de colesterol. Os objetivos específicos são:

- Construção de modelos teóricos das bicamadas lipídicas com diferentes concentrações de colesterol;
- Simulações por DM das bicamadas lipídicas;
- Avaliação dos parâmetros estruturais e físico-químicos das bicamadas lipídicas simuladas;
- Avaliação do comportamento das bicamadas lipídicas frente a diferentes concentrações de colesterol.

3 METODOLOGIA

3.1 Construção das bicamadas lipídicas

As estruturas das bicamadas lipídicas contendo diferentes tipos de lipídeos foram construídas através do servidor CHARMM-GUI^{59,60}. Foram utilizadas moléculas de POPC, DOPE, POPS e colesterol para construção das bicamadas lipídicas (Figura 3). Na Tabela 1 estão dispostas as diferentes bicamadas lipídicas que foram construídas, bem como o número de lipídeos em cada uma delas, o número de moléculas de solvente e dos íons Na⁺ e Cl⁻. O número total de cada um dos tipos de lipídeos foi dividido igualmente entre as duas monocamadas de cada bicamada lipídica.

Figura 3 - Representação estrutural dos lipídeos utilizados nas simulações de bicamadas lipídicas.

Tabela 1 - Modelos de bicamadas lipídicas construídos no servidor CHARMM-GUI^{59,60} e suas respectivas composições em termos do número de moléculas de lipídeos, água e íons.

Sistema	POPC	DOPE	POPS	Colesterol	Solvente	Na ⁺	Cl ⁻
Controle	414	-	-	-	20700	49	49
1	214	166	10	62	22600	59	49
2	178	166	10	122	23800	62	52
3	130	166	10	206	25600	67	57

O número de moléculas de colesterol presentes nas bicamadas lipídicas foi determinado com base em composições lipídicas experimentais da camada externa de membranas sinápticas. A mesma composição da camada externa foi utilizada também na camada interna das bicamadas lipídicas construídas. Para o sistema 1, por exemplo, foi construída uma bicamada lipídica com 32 moléculas de colesterol por monocamada. Portanto, cada monocamada terá concentração de colesterol correspondente à camada externa de uma membrana sináptica com distribuição de colesterol de 15%/85% (camada externa/camada interna) em relação a sua concentração total (18% em massa). Da mesma forma, foram determinadas as concentrações de colesterol para os sistemas 2 e 3 que correspondem, respectivamente, à camada externa de uma membrana sináptica com distribuição de colesterol de 30%/70% e 50%/50% (camada externa/camada interna) em relação a sua concentração total (18% em massa).

Os íons Na^+ e Cl^- foram adicionados aos sistemas para neutralizá-los, mantendo uma concentração de 0,15 M. O número de moléculas de água foi determinado de forma que para cada lipídeo fossem adicionadas em torno de 50 moléculas. Moléculas de água do tipo TIP3P foram utilizadas e os sistemas foram inseridos em caixas de simulação do tipo retangular.

3.2 Simulações por DM

Para realização das simulações por DM, o campo de forças CHARMM36⁵⁴ e o programa NAMD⁵² foram utilizados. As etapas de equilibração foram simuladas utilizando-se o *ensemble* NVT (primeiros 50 ps) seguido do *ensemble* NPdT (325 ps restantes). A fase de equilibração foi dividida em seis etapas, conforme sugerido pelo módulo *Membrane Builder* do servidor CHARMM-GUI^{61,62}, sendo que as constantes de força para os componentes do sistema foram sendo reduzidas ao longo dessas etapas (Tabela 2).

Tabela 2 - Etapas de equilibração das simulações por DM dos sistemas investigados.

Etapa	Ensemble	Tempo de integração	Tempo de equilibração	Constantes de força para restrições harmônicas ($\text{kcal mol}^{-1} \text{\AA}^{-2}$)		
				Água ¹	Lipídeo ²	Íon ³
1	NVT	1 fs	25 ps	2,5	2,5	10,0
2	NVT	1 fs	25 ps	2,5	2,5	0,0
3	NPdT	1 fs	25 ps	1,0	1,0	0,0
4	NPdT	2 fs	100 ps	0,5	0,5	0,0

Etapa	Ensemble	Tempo de integração	Tempo de equilíbrio	Constantes de força para restrições harmônicas (kcal mol ⁻¹ Å ⁻²)		
				Água ¹	Lípídeo ²	Íon ³
5	NPaT	2 fs	100 ps	0,1	0,1	0,0
6	NPaT	2 fs	100 ps	0,0	0,0	0,0

¹Restrições harmônicas para manter as moléculas de água longe da região hidrofóbica da membrana. ²Restrições harmônicas para manter a cauda dos lipídeos em $-5 \text{ Å} < Z < 5 \text{ Å}$ e a cabeça dos lipídeos próximas à superfície da membrana. ³Restrições harmônicas de posição.

O *ensemble* NPaT também foi empregado nas etapas de produção. O tempo de integração para a etapa de produção foi definido em 2 fs. O sistema controle foi simulado a temperaturas de 298, 310 e 323 K, enquanto que os sistemas 1, 2 e 3 foram simulados à temperatura de 310 K, todos com condições periódicas de contorno. Para garantir uma temperatura constante, utilizou-se a dinâmica de Langevin. Todos os sistemas foram simulados à pressão de 1 bar, utilizando-se o método *Nosé-Hoover Langevin Piston* para o controle da pressão. Para as interações eletrostáticas e de van der Waals entre termos não ligados foi utilizado um raio de corte de 12 Å e função de suavização a partir de 10 Å. Para as interações eletrostáticas, foi utilizado o método *Particle-Mesh Ewald* (PME).

Para o sistema controle foram simulados 50 ns nas etapas de produção para as três diferentes temperaturas: 298, 310 e 323 K. Após uma análise dos resultados obtidos e validação da metodologia empregada, simulações de 300 ns foram realizadas apenas na temperatura de interesse biológico de 310 K para todos os sistemas. O sistema controle, contendo apenas moléculas de POPC, solvente e íons, foi utilizado como um sistema de referência nesse estudo.

As propriedades estruturais dos sistemas simulados por DM molecular foram determinadas utilizando-se programas implementados no pacote GROMACS⁵¹ e o programa VMD⁶³. Avaliou-se o comportamento das bicamadas lipídicas em relação aos valores de APL, EB, DL, PO, perfis de densidade e análises visuais da trajetória obtida.

4 RESULTADOS E DISCUSSÃO

4.1 Construção dos sistemas

De acordo com os resultados experimentais, 18% (em massa) das membranas sinápticas corresponde ao colesterol, sendo que há uma variação na sua distribuição entre as camadas dessas membranas.^{12,15} Nesse estudo foram investigadas bicamadas lipídicas cuja distribuição de colesterol se dá da seguinte forma entre as diferentes camadas (camada externa/camada interna): 15%/85%; 30%/70% e 50%/50%.

Inicialmente, foram realizadas algumas simulações de DM na tentativa de se utilizar diferentes composições de lipídeos em cada monocamada da bicamada lipídica, de forma a se construir modelos mais compatíveis com os sistemas biológicos. Entretanto, os resultados mostraram uma instabilidade dos sistemas simulados, além de apresentarem redução significativa da APL durante todo o tempo de simulação. Estudos desenvolvidos por Park e colaboradores (2015) já têm demonstrado essa dificuldade em simular bicamadas lipídicas com diferentes composições de lipídeos entre as monocamadas.⁵⁸

Frente aos resultados obtidos, optou-se por construir sistemas em que a composição de lipídeos fosse a mesma para as duas monocamadas das bicamadas lipídicas. Uma vez que o interesse maior em relação às membranas sinápticas está nas interações que ocorrem com sua superfície externa, foram simuladas bicamadas lipídicas em que a distribuição de lipídeos nas duas camadas se aproximasse da composição experimental da superfície externa da membrana. Dessa forma, as duas monocamadas mantiveram a mesma composição de lipídeos, representando, portanto, a camada externa de uma membrana sináptica.

4.2 Influência da temperatura nas propriedades estruturais do sistema controle

Para se realizar os estudos com as membranas sinápticas, fez-se necessário utilizar um sistema controle, neste caso bicamadas lipídicas compostas apenas por moléculas de POPC. As propriedades estruturais obtidas a partir das simulações por DM das bicamadas lipídicas de POPC foram comparadas com as mesmas propriedades disponíveis na literatura e que foram determinadas experimentalmente.

Inicialmente, avaliaram-se os valores de APL das etapas de produção do sistema controle em três diferentes temperaturas (298, 310 e 323 K). Comparando-se os valores de

APL das simulações nas três temperaturas, os resultados mostram um aumento dos valores de APL em função do aumento da temperatura de simulação (Figura 4). Kucerka e colaboradores (2011) já têm demonstrado que um aumento de temperatura acarreta em alterações nos parâmetros estruturais de bicamadas lipídicas compostas por fosfatidilcolinas, com consequente aumento da APL e diminuição da EB.⁴³

Figura 4 - APL para o sistema controle em diferentes temperaturas.

Os valores médios de APL e EB (após 20 ns de simulação) foram comparados aos valores experimentais determinados por Kucerka e colaboradores (2011).⁴³ A partir de análise de regressão linear aplicada aos valores experimentais (Apêndices A e B), obtiveram-se valores estimados de APL e EB, respectivamente, para as temperaturas de interesse. Os valores calculados pelas simulações por DM foram então comparados aos valores estimados baseando-se nos resultados experimentais (Tabela 3).

Tabela 3 - Valores calculados (*Cal.*) e experimentais (*Exp.*) de APL e EB para o sistema controle.

Temp. (K)	APL			EB		
	<i>Cal.</i> (Å ²)	<i>Exp.</i> (Å ²) ^a	<i>Resíduo Relativo</i> (%) ^c	<i>Cal.</i> (Å)	<i>Exp.</i> (Å) ^b	<i>Resíduo Relativo</i> (%) ^c
298	63,7	63,5	0,3	39,1	37,1	5,4
310	65,3	65,1	0,3	38,8	36,2	7,2
323	66,9	67,0	0,1	38,5	35,2	9,4

^aValores determinados a partir de regressão linear aplicada aos valores experimentais⁴³ (Apêndice A). ^bValores determinados a partir de regressão linear aplicada aos valores experimentais⁴³ (Apêndice B). ^cOs resíduos relativos são obtidos dividindo-se os valores dos resíduos absolutos (em módulo) pelos valores de APL experimental correspondente e multiplicando-se o resultado por 100.

Os valores de APL determinados a partir das simulações por DM variam de 63,7 a 66,9 Å², em função do aumento da temperatura, enquanto que os valores experimentais variam de 63,5 a 67,0 Å². Uma análise do resíduo relativo mostra que os resultados teóricos se aproximam dos resultados experimentais, já que o maior desvio observado é de apenas 0,3%. Portanto, a metodologia empregada foi capaz de determinar satisfatoriamente os valores de APL para o sistema controle.

Analisando-se os valores de EB para o sistema controle, pode-se observar o mesmo comportamento tanto para os resultados experimentais quanto para os teóricos. Os valores de EB diminuem em função do aumento da temperatura, variando de 39,1 a 38,5 para as simulações por DM e de 37,1 a 35,2 para os resultados experimentais. Neste caso, os desvios entre os valores calculados e experimentais chegam a 9,4%.

Na Figura 5 é possível observar dois picos referentes à densidade do grupo fosfato das moléculas de POPC em relação à distância do centro da bicamada lipídica na simulação por DM a 310 K. O cálculo dos valores de EB foi feito tomando-se a distância entre esses dois picos, para cada simulação realizada.

Figura 5 - Densidade do grupo fosfato das moléculas de POPC em relação ao centro do sistema controle na simulação realizada a 310 K.

Também foi avaliada a DL para o sistema controle simulado nas três diferentes temperaturas (Tabela 4). Os resultados obtidos a partir das simulações por DM são comparados aos resultados experimentais determinados por Filippov e colaboradores (2003).⁶⁴

Tabela 4 - Valores calculados (*Cal.*) e experimentais (*Exp.*) de DL para o sistema controle.

Temp. (K)	DL ($10^{-12} \text{ m}^2 \text{ s}^{-1}$)		
	<i>Cal.</i>	<i>Exp.</i>	<i>Resíduo Relativo (%)</i> ^b
298	8,5	8,9	4,5
310	13,3	14,3 ^a	7,0
323	15,1	20,6	26,7

^aValor determinado a partir de regressão linear aplicada aos valores experimentais⁶⁴ (Apêndice C). ^bOs resíduos relativos são obtidos dividindo-se os valores dos resíduos absolutos (em módulo) pelos valores de APL experimental correspondente e multiplicando-se o resultado por 100.

Os resultados mostram uma boa concordância entre valores experimentais e teóricos para as temperaturas de 298 e 310 K. Entretanto, para temperaturas elevadas como 323 K, há uma diferença significativa entre os valores experimental e calculado. De forma geral, a DL do sistema controle cresce com o aumento da temperatura (de $8,5 \times 10^{-12}$ a $15,1 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$), assim como foi observado experimentalmente por Filippov e colaboradores (2003).⁶⁴ A determinação dos valores de DL é feita a partir do deslocamento quadrático médio (DQM) dos átomos em relação a uma posição de referência. A Figura 6 mostra que os átomos das moléculas de POPC têm um deslocamento maior em altas temperaturas.

Figura 6 - DQM das moléculas de POPC no sistema controle simulado nas temperaturas 298, 310 e 323 K.

Na Figura 7 pode-se observar a variação do PO (representado pelo $|S_{CH}|$) para o sistema controle nas diferentes temperaturas. Um aumento na temperatura provoca uma diminuição nos valores de PO, o que corresponde a um sistema menos ordenado, ou seja, menor é o alinhamento das cadeias alifáticas. Seelig e Waespe-Sarcevic (1978) e Perly e

colaboradores (1985) determinaram experimentalmente os valores de PO para alguns carbonos das cadeias acil de bicamadas lipídicas compostas por moléculas de POPC na temperatura de 300 K. Os resultados obtidos experimentalmente também estão dispostos na Figura 7.^{65,66}

Figura 7 - POs obtidos a partir das simulações por DM do sistema controle em função da variação da temperatura.

Analisando-se a Figura 7 é possível observar que os átomos de carbono mais próximos das extremidades das cadeias acil não sofrem tanta influência da temperatura quanto os átomos mais internos. Para os átomos C2 e C16 da cadeia *sn1* e os átomos C2 e C18 da cadeia *sn2* das moléculas de POPC, os valores de PO são semelhantes, independente da temperatura de simulação.

Os resultados mostram ainda que os valores de PO obtidos a partir das simulações por DM apresentam comportamento semelhante aos valores determinados experimentalmente. Para a cadeia *sn2* das moléculas de POPC, por exemplo, observa-se uma redução dos valores de PO para os átomos envolvidos na ligação dupla (átomos C9 e C10) e também para os átomos a eles ligados (C8 e C11), em comparação aos átomos C3 a C7 e C12 a C16. Esse comportamento é característico das cadeias dos fosfolipídeos que apresentam insaturações.

Muitas propriedades estruturais e dinâmicas das bicamadas lipídicas são governadas pela temperatura, tais como: APL, EB, DL e PO.²² A partir das simulações por DM do sistema controle em diferentes temperaturas, foi possível observar que os valores de APL e DL aumentam em função da temperatura, enquanto que há redução nos valores de EB e PO,

conforme determinado experimentalmente. Em temperaturas mais elevadas, as bicamadas lipídicas adquirem maior fluidez, o que provoca alterações significativas em seus parâmetros estruturais. Consequentemente, as cadeias acil dos fosfolipídeos se tornam mais desordenadas, o que provoca uma diminuição dos valores de PO e EB. Além disso, os fosfolipídeos apresentam maior mobilidade, acarretando em um aumento da APL e da DL.

As simulações por DM do sistema controle nas temperaturas 298, 310 e 323 K mostram boa concordância com os resultados experimentais, validando assim a metodologia empregada. Portanto, serão estudados os demais sistemas onde há variações da concentração de colesterol utilizando a mesma metodologia empregada para o sistema controle.

4.3 Influência do colesterol nas propriedades estruturais dos sistemas 1, 2 e 3

Após as simulações iniciais para identificação do comportamento dos sistemas em diferentes temperaturas e avaliação da metodologia empregada, a simulação por DM do sistema controle na temperatura de 310 K foi estendida até 300 ns, sendo que também foram simulados 300 ns para os sistemas 1 a 3, à mesma temperatura. A Figura 8 apresenta o sistema controle, simulado por DM.

Figura 8 - Representação do sistema controle simulado a 310 K no tempo de simulação de 300 ns. As esferas vermelhas representam os íons de sódio e as esferas verdes os íons de cloro.

Uma análise visual das trajetórias dos sistemas 1, 2 e 3 demonstram a distribuição dos lipídeos nas bicamadas lipídicas, bem como o deslocamento lateral dos mesmos ao longo das simulações (Figura 9). Os sistemas possuem o mesmo número de moléculas de DOPE e

POPS, enquanto que há variação na concentração de moléculas de colesterol e de POPC, conforme apresentado na Tabela 1.

Figura 9 - Representação dos sistemas 1, 2 e 3 simulados a 310 K nos tempos de simulação de 0, 100, 200 e 300 ns. As moléculas de POPC, DOPE, POPS e colesterol estão representadas pelas cores azul, amarelo, vermelho e verde, respectivamente.

Para que os valores de APL dos sistemas 1, 2 e 3 fossem comparados aos valores obtidos para o sistema controle, e uma vez que não há dados experimentais disponíveis para membranas com as mesmas composições desses sistemas, os valores de APL foram determinados dividindo-se a área da bicamada pelo número de moléculas de lipídeos presentes em cada monocamada, incluindo as moléculas de colesterol. Dessa forma, será possível avaliar o comportamento das bicamadas lipídicas frente à variação da concentração de moléculas de colesterol.

A partir das simulações por DM dos sistemas investigados, obtiveram-se os valores de APL em função do tempo de simulação, conforme demonstrado na Figura 10. Os resultados

indicam que o aumento na concentração de colesterol na bicamada provoca diminuição da APL.

Figura 10 - APLs obtidas a partir das simulações por DM do sistema controle e dos sistemas 1, 2 e 3.

A variação dos valores de APL para os sistemas investigados evidencia a estabilidade dos mesmos durante as simulações por DM. Além disso, os resultados mostram uma APL média de $65,4 \text{ \AA}^2$ para o sistema controle (após 20 ns de simulação). Conforme descrito anteriormente, o valor experimental para essa propriedade determinado a partir dos valores experimentais obtidos por Kucerca e colaboradores (2011) foi de $65,1 \text{ \AA}^2$ (Apêndice A), o que corresponde a um desvio relativo de 0,5% entre o valor experimental e o valor calculado a partir das simulações por DM.⁴³

Para os sistemas 1 a 3, os valores médios de APL, após 20 ns de simulação foram de $56,3$, $51,2$, e $46,3 \text{ \AA}^2$, respectivamente. Entretanto, na literatura não há valores experimentais para bicamadas lipídicas com as mesmas composições aqui utilizadas. Portanto, os valores de APL obtidos pelas simulações por DM não devem ser usados como valores absolutos, e sim como indicativos da estabilidade das simulações e da tendência em relação à variação da concentração de colesterol.

Interações entre as moléculas de colesterol e os fosfolipídeos são capazes de afetar a APL da bicamada lipídica. Como consequência, há diminuição da APL em função do aumento da concentração de colesterol conforme observado a partir das simulações por DM e de resultados experimentais. Esse comportamento é um indicativo do efeito de condensação

provocado pelas moléculas de colesterol, produzindo assim uma bicamada lipídica mais compacta.

Avaliando-se os valores de PO (representado pelo $|S_{CH}|$) para os sistemas estudados, é possível observar que o aumento na concentração de colesterol induz uma maior ordenação das cadeias alifáticas (Figura 11).

Figura 11 - POs obtidos a partir das simulações por DM do sistema controle e dos sistemas 1, 2 e 3.

Os resultados mostram um ordenamento das cadeias alifáticas das moléculas de POPC, DOPE e POPS em função do aumento da concentração de colesterol. As moléculas de colesterol, ao estabelecerem interações com as cadeias acil dos fosfolipídeos, provocam um maior ordenamento das mesmas. De forma geral, esse efeito é menos pronunciado nas extremidades das cadeias, uma vez que elas apresentam menor contato com as moléculas de colesterol, conforme ilustrado pela Figura 12.

Figura 12 - Posição da molécula de colesterol em relação a moléculas de POPC no sistema 1. A molécula de colesterol está na representação de esferas de van der Waals.

Avaliando-se os valores de PO para as moléculas de POPC no sistema controle em comparação com os demais sistemas, é possível observar uma diferença significativa entre esses valores. Para os átomos de carbono que participam da ligação dupla da cadeia *sn2* das moléculas de POPC, por exemplo, os valores de PO do sistema controle são 0,060 e 0,046 para os átomos C9 e C10, respectivamente. Já para o sistema 3, que tem a maior concentração de colesterol dentre os sistemas estudados, esses valores aumentam para 0,123 e 0,103. Ferreira e colaboradores (2013) investigaram o efeito do colesterol em bicamadas lipídicas de POPC a partir de experimentos de ressonância magnética nuclear e simulações por DM. Eles variaram a concentração de colesterol de 0 a 60% e observaram o mesmo efeito para o ordenamento dos átomos de carbono das cadeias acil das moléculas de POPC, ou seja, em concentrações mais altas de colesterol, o sistema se mostra mais ordenado.⁶⁷

Para as moléculas de DOPE e POPS, observa-se o mesmo comportamento das moléculas de POPC em relação aos valores de PO. Quanto maior a concentração de colesterol

na bicamada lipídica, maior são os valores de PO, principalmente para os átomos de carbono mais internos das cadeias acil. Para as moléculas de DOPE, por exemplo, os valores de PO para os átomos C9 e C10 da cadeia *sn1* são 0,109 e 0,092 (sistema 1), 0,127 e 0,104 (sistema 2) e 0,148 e 0,114 (sistema 3) e para os mesmos átomos da cadeia *sn2*, os valores são de 0,085 e 0,075 (sistema 1), 0,103 e 0,088 (sistema 2) e 0,123 e 0,103 (sistema 3). Já os valores de PO para os átomos C9 e C10 da cadeia *sn2* das moléculas de POPS variam de 0,094 e 0,070 a 0,119 e 0,100, do sistema com menor concentração de colesterol ao sistema com maior concentração.

A redução da APL e o aumento do PO em função de um aumento da concentração de colesterol também foram observados por Hofsäß e colaboradores (2003) ao simularem bicamadas lipídicas com DPPC e colesterol. Os autores demonstram ainda que esses dois parâmetros estruturais estão correlacionados entre si.⁴⁶

Avaliaram-se também os valores de EB para os sistemas estudados. Os resultados obtidos a partir das simulações por DM permitem inferir que a presença de moléculas de colesterol provoca aumento dos valores de EB (Tabela 5). Contudo, para o sistema 3, com maior concentração de colesterol, esse feito já não se torna tão pronunciado, tal como observado entre os sistemas 1 e 2. Leftin e colaboradores (2014) também observaram que a presença de colesterol em bicamadas lipídicas é capaz de aumentar significativamente a espessura de bicamadas compostas por POPC e colesterol.²⁷

Tabela 5 - EB para os sistemas simulados a 310 K.

Sistema		EB (Å)
Controle	POPC	38,3
	POPC	42,9
1	DOPE	39,2
	POPS	42,9
	POPC	46,0
	DOPE	42,0
2	POPS	46,0
	POPC	46,0
	DOPE	46,0
3	POPS	46,0
	DOPE	46,0
	POPC	46,0

O valor de EB para o sistema controle determinado a partir da simulação por DM (38,3 Å) está próximo do valor obtido experimentalmente (36,2 Å; Apêndice B), com um desvio relativo de 5,8%.⁴³ Comparando-se esses valores de EB para as moléculas de POPC com aqueles obtidos nos sistemas que contêm moléculas de colesterol, observa-se que um aumento significativo. Para o sistema com maior concentração de colesterol, o valor de EB para as moléculas de POPC chega a 46 Å. O mesmo comportamento é observado também para as moléculas de DOPE e POPS.

Na Figura 13 estão representados os perfis de densidade ao longo da bicamada lipídica para os componentes dos sistemas estudados (POPC, DOPE, POPS, colesterol, solvente). Os resultados obtidos mostram concordância com os valores de EB, uma vez que à medida que se aumenta a concentração de colesterol, a densidade dos fosfolipídeos abrange uma região maior ao longo do eixo z.

Figura 13 - Perfis de densidade obtidos a partir das simulações por DM do sistema controle e dos sistemas 1, 2 e 3.

Em relação à DL, foi possível observar uma diminuição acentuada dos valores quando se aumenta a concentração de colesterol (Tabela 6). No apêndice D estão dispostas as variações dos valores de deslocamento quadrático médio (DQM) em função do tempo de simulação para os sistemas 1, 2 e 3 e também para o sistema controle.

Tabela 6 - DL para os sistemas simulados a 310 K.

Sistema		DL ($10^{-12} \text{ m}^2 \text{ s}^{-1}$)
Controle	<i>POPC</i>	9,6
	<i>POPC</i>	7,3
1	<i>DOPE</i>	7,2
	<i>POPS</i>	10,2
	<i>POPC</i>	6,0
2	<i>DOPE</i>	5,1
	<i>POPS</i>	3,6
	<i>POPC</i>	3,1
3	<i>DOPE</i>	3,0
	<i>POPS</i>	4,6

As interações que acontecem entre o colesterol e as cadeias acil dos fosfolipídeos possivelmente impedem a mobilidade dos mesmos na direção perpendicular à bicamada lipídica, o que provoca uma redução significativa dos valores de DL. Para as moléculas de POPC no sistema controle, o valor de DL para a simulação de 300 ns a 310 K foi de $9,6 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$, enquanto que na simulação mais curta (50 ns), discutida anteriormente, esse valor foi de $13,3 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$. Acredita-se que essa discrepância nos valores de DL seja causada pelo efeito do tempo de simulação, já observado em estudos com bicamadas lipídicas de DPPC, onde os valores de DL são menores para simulações mais longas em comparação a simulações mais curtas.^{44,45}

Comparando-se os valores de DL para as moléculas de POPC nos diferentes sistemas, observa-se uma redução que vai de $9,6 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$ no sistema controle para $3,1 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$ no sistema 3. Os sistemas 1 e 2 apresentam valores de DL para as moléculas de POPC de $7,3 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$ e $6,0 \times 10^{-12} \text{ m}^2 \text{ s}^{-1}$, respectivamente. De forma geral, a redução dos valores de DL em função do aumento da concentração de colesterol é observado também para as moléculas de DOPE e POPS.

Os resultados obtidos a partir das simulações por DM mostram que a concentração de colesterol em uma bicamada lipídica afeta de forma significativa suas propriedades estruturais. As moléculas de colesterol presentes em uma bicamada lipídica são capazes de interagir com os fosfolípidos em função de efeitos hidrofóbicos e interações de van der Waals. Essas interações podem ocorrer entre as cadeias acil dos fosfolípidos e a estrutura planar e rígida da molécula de colesterol.²⁴⁻²⁶

Portanto, quanto maior a concentração de colesterol em uma bicamada lipídica, maior será o ordenamento das cadeias acil dos fosfolípidos. Como consequência, observa-se um aumento dos valores de PO e redução da DL em função do aumento da concentração de colesterol. Os resultados experimentais e os resultados obtidos nesse trabalho mostram ainda a condensação da APL e o aumento da EB em função do aumento da concentração de colesterol.

As simulações por DM dos sistemas controle, 1, 2 e 3 apresentaram resultados satisfatórios e coerentes com observações experimentais. Os resultados aqui obtidos demonstram que as bicamadas lipídicas foram adequadamente simuladas e que estão também aptas para serem utilizadas em simulações posteriores.

5 CONCLUSÕES E PERSPECTIVAS

As simulações por DM desenvolvidas neste trabalho permitiram propor modelos teóricos de bicamadas lipídicas que se assemelham à composição lipídica de membranas sinápticas. A metodologia empregada foi validada a partir de simulações por DM de bicamadas lipídicas compostas por moléculas de POPC em três diferentes temperaturas. Após as simulações iniciais, três sistemas foram propostos para representar membranas sinápticas com diferentes concentrações de colesterol.

De forma geral, as simulações evidenciaram os efeitos provocados pela presença de moléculas de colesterol em bicamadas lipídicas. Em função das interações que acontecem entre as moléculas de colesterol e os fosfolipídeos, principalmente as cadeias acil, as propriedades estruturais das bicamadas são afetadas. Observou-se que quanto maior a concentração de colesterol, maiores são as alterações dessas propriedades.

Foram avaliados nesse trabalho os valores de APL, EB, DL e PO para os sistemas 1, 2 e 3 e o sistema controle, que diferenciam entre si em relação à concentração de moléculas de colesterol. Observou-se que o aumento na concentração de colesterol induz redução dos valores de APL e DL e aumento dos valores de EB e PO. Em concentrações mais elevadas, as moléculas de colesterol são capazes de promover um ordenamento maior das cadeias acil dos fosfolipídeos, o que também contribui para a alteração das propriedades estruturais investigadas.

O estudo desenvolvido permitiu propor modelos mais realísticos para simulações de membranas sinápticas que serão disponibilizados em breve para acesso público. Com isso, contribui-se para o desenvolvimento de novas pesquisas que buscam compreender mecanismos associados a interações entre receptores e lipídeos ou mesmo entre pequenas moléculas como os neurotransmissores e a membrana sináptica.

Em função dos resultados obtidos, os sistemas aqui propostos já estão sendo utilizados para outros estudos envolvendo receptores e neurotransmissores, tais como o receptor muscarínico e moléculas de acetilcolina. Além disso, a mesma metodologia está sendo aplicada para a equilibração de bicamadas com concentração de lipídeos que se assemelham a membranas mitocondriais.

REFERÊNCIAS

1. NELSON, D. L.; COX, M. M. **Lehninger principles of biochemistry**. 6. ed. New York: W. H. Freeman, 2013. 1198 p.

2. BITOUNIS, D.; FANCIULLINO, R.; ILIADIS, A.; CICCOLINI, J. Optimizing druggability through liposomal formulations: new approaches to an old concept. **ISRN Pharmaceutics**, v. 2012, p. 738432, 2012.

3. SINGER, S. J.; NICOLSON, G. L. The fluid mosaic model of the structure of cell membranas. **Science**, v. 175, n. 4023, p. 720–731, 1972.

4. SIMONS, K.; SAMPAIO, J. L. Membrane organization and lipid rafts. **Cold Spring Harbor Perspectives in Biology**, v. 3, n. 10, p. a004697, 2011.

5. NICOLSON, G. L. The fluid-mosaic model of membrane structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. **Biochimica et Biophysica Acta**, v. 1838, n. 6, p. 1451–1466, 2014.

6. PÖYRY, S.; CRAMARIUC, O.; POSTILA, P. A.; KASZUBA, K.; SAREWICZ, M.; OSYCZKA, A.; VATTULAINEN, I.; RÓG, T. Atomistic simulations indicate cardiolipin to have an integral role in the structure of the cytochrome bc1 complex. **Biochimica et Biophysica Acta**, v. 1827, n. 6, p. 769–778, 2013.

7. KOPEĆ, W.; TELENIS, J.; KHANDELIA, H. Molecular dynamics simulations of the interactions of medicinal plant extracts and drugs with lipid bilayer membranes. **The FEBS Journal**, v. 280, n. 12, p. 2785–2805, 2013.

8. VAUQUELIN, G.; PACKEU, A. Ligands, their receptors and ... plasma membranes. **Molecular and Cellular Endocrinology**, v. 311, n. 1-2, p. 1–10, 2009.

9. WANG, D.; GOU, S.-Y.; AXELROD, D. Reaction rate enhancement by surface diffusion of adsorbates. **Biophysical Chemistry**, v. 43, n. 2, p. 117–137, 1992.

10. POSTILA, P. A.; VATTULAINEN, I.; RÓG, T. Selective effect of cell membrane on synaptic neurotransmission. **Scientific Reports**, v. 6, p. 19345, 2016.

11. HAN, X. Multi-dimensional mass spectrometry-based shotgun lipidomics and the altered lipids at the mild cognitive impairment stage of Alzheimer's disease. **Biochimica et Biophysica Acta**, v. 1801, n. 8, p. 774–783, 2010.
12. COTMAN, C. W.; BLANK, M. L.; MOEHL, A.; SNYDER, F. Lipid composition of synaptic plasma membranes isolated from rat brain by zonal centrifugation. **Biochemistry**, v. 8, n. 11, p. 4606–4612, 1969.
13. WOOD, W. G.; IGBAVBOA, U.; MÜLLER, W. E.; ECKERT, G. P. Cholesterol asymmetry in synaptic plasma membranes. **Journal of Neurochemistry**, v. 116, n. 5, p. 684–689, 2011.
14. SVENNERHOLM, L.; BOSTROM, K.; HELANDER, C. G.; JUNGBJER, B. Membrane lipids in the aging human brain. **Journal of Neurochemistry**, v. 56, n. 6, p. 2051–2059, 1991.
15. IGBAVBOA, U.; AVDULOV, N. A.; SCHROEDER, F.; WOOD, W. G. Increasing age alters transbilayer fluidity and cholesterol asymmetry in synaptic plasma membranes of mice. **Journal of Neurochemistry**, v. 66, n. 4, p. 1717–1725, 2002.
16. SHENG, R.; CHEN, Y.; YUNG GEE, H.; STEC, E.; MELOWIC, H. R.; BLATNER, N. R.; TUN, M. P.; KIM, Y.; KÄLLBERG, M.; FUJIWARA, T. K.; HYE HONG, J.; PYO KIM, K.; LU, H.; KUSUMI, A.; GOO LEE, M.; CHO, W. Cholesterol modulates cell signaling and protein networking by specifically interacting with PDZ domain-containing scaffold proteins. **Nature Communications**, v. 3, p. 1249, 2012.
17. SCHENGRUND, C.-L. Lipid rafts: keys to neurodegeneration. **Brain Research Bulletin**, v. 82, n. 1–2, p. 7–17, 2010.
18. KAISER, H.-J.; ORLOWSKI, A.; RÓG, T.; NYHOLM, T. K. M.; CHAI, W.; FEIZI, T.; LINGWOOD, D.; VATTULAINEN, I.; SIMONS, K. Lateral sorting in model membranes by cholesterol-mediated hydrophobic matching. **Proceeding of the National Academy of Sciences of the United States of America**, v. 108, n. 40, p. 16628–16633, 2011.
19. LEVITAN, I.; SINGH, D. K.; ROSENHOUSE-DANTSKEER, A. Cholesterol binding to ion channels. **Frontiers in Physiology**, v. 5, p. 65, 2014.
20. PUCADYIL, T. J.; CHATTOPADHYAY, A. Role of cholesterol in the function and organization of G-protein coupled receptors. **Progress in Lipid Research**, v. 45, n. 4, p. 295–333, 2006.

21. HONG, W. C.; AMARA, S. G. Membrane cholesterol modulates the outward facing conformation of the dopamine transporter and alters cocaine binding. **The Journal of Biological Chemistry**, v. 285, n. 42, p. 32616–32626, 2010.

22. NAGLE, J. F.; TRISTRAM-NAGLE, S. Structure of lipid bilayers. **Biochimica et Biophysica Acta**, v. 1469, n. 3, p. 159–195, 2000.

23. AVANTI POLAR LIPIDS. 2016. Disponível em: <<http://www.avantilipids.com>>. Acesso em: 15 jan. 2016.

24. MARTINEZ-SEARA, H.; RÓG, T.; KARTTUNEN, M.; VATTULAINEN, I.; REIGADA, R. Cholesterol induces specific spatial and orientational order in cholesterol/phospholipid membranes. **PLoS ONE**, v. 5, v. 6, p. e11162, 2010.

25. MITCHELL, D. C.; LITMAN, B. J. Effect of cholesterol on molecular order and dynamics in highly polyunsaturated phospholipid bilayers. **Biophysical Journal**, v. 75, n. 2, p. 896–908, 1998.

26. STOCKTON, G. W.; SMITH, I. C. P. A deuterium nuclear magnetic resonance study of the condensing effect of cholesterol on egg phosphatidylcholine bilayer membranes. I. Perdeuterated fatty acid probes. **Chemistry and Physics of Lipids**, v. 17, n. 2-3, p. 251–263, 1976.

27. LEFTIN, A.; MOLUGU, T. R.; JOB, C.; et al. (2014) Area per lipid and cholesterol interactions in membranes from separated local-field ^{13}C NMR spectroscopy. **Biophysical Journal**, v. 107, n. 10, p. 2274–2286, 2014.

28. MCINTOSH, T. J. The effect of cholesterol on the structure of phosphatidylcholine bilayers. **Biochimica et Biophysica Acta**, v. 513, n. 1, p. 43–58, 1978.

29. FIORINI, R.; GRATTON, E.; CURATOLA, G. Effect of cholesterol on membrane microheterogeneity: a study using 1,6-diphenyl-1,3,5-hexatriene fluorescence lifetime distributions. **Biochimica et Biophysica Acta**, v. 1006, n. 2, p. 198–202, 1989.

30. DE MEYER, F. J.-M.; RODGERS J. M.; WILLEMS, T. F.; SMIT, B. Molecular simulation of the effect of cholesterol on lipid-mediated protein-protein interactions. **Biophysical Journal**, v. 99, n. 11, p. 3629–3638, 2010.

31. ANDERSON, T. G.; MCCONNELL, H. M. Condensed complexes and the calorimetry of cholesterol-phospholipid bilayers. **Biophysical Journal**, v. 81, n. 5, p. 2774–2785, 2001.

32. RADHAKRISHNAN, A.; LI, X. M.; BROWN, R. E.; MCCONNELL, H. M. Stoichiometry of cholesterol-sphingomyelin condensed complexes in monolayers. **Biochimica et Biophysica Acta**, v. 1511, n. 1, p. 1–6, 2001.
33. SIMONS, K.; EHEHALT, R. Cholesterol, lipid rafts, and disease. **The Journal of Clinical Investigation**, v. 110, n. 5, p. 597–603, 2002.
34. ALMEIDA, P. F. F. Thermodynamics of lipid interactions in complex bilayers. **Biochimica et Biophysica Acta**, v. 1788, n. 1, p. 72–85, 2009.
35. IPSEN, J. H.; KARLSTRÖM, G.; MOURTISEN, O. G.; WENNERSTRÖM, H.; ZUCKERMANN, M. J. Phase equilibria in the phosphatidylcholine-cholesterol system. **Biochimica et Biophysica Acta**, v. 905, n. 1, p. 162–172, 1987.
36. WOOLF, T. B.; ROUX, B. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 24, p. 11631–11635, 1994.
37. EDHOLM, O.; BERGER, O.; JÄHNIG, F. Structure and fluctuations of bacteriorhodopsin in the purple membrane: a molecular dynamics study. **Journal of Molecular Biology**, v. 250, n. 1, p. 94–111, 1995.
38. LIGUORI, N.; NERENBERG, P. S.; HEAD-GORDON, T. Embedding A β 42 in heterogeneous membranes depends on cholesterol asymmetries. **Biophysical Journal**, v. 105, n. 4, p. 899–910, 2013.
39. CAO, R.; ROSSETTI, G.; BAUER, A.; CARLONI, P. Binding of the antagonist caffeine to the human adenosine receptor hA2AR in nearly physiological conditions. **PLoS ONE**, v. 10, n. 5, p. e0126833, 2015.
40. MONJE-GALVAN, V.; KLAUDA, J. B. Modeling yeast organelle membranes and how lipid diversity influences bilayer properties. **Biochemistry**, v. 54, n. 45, p. 6852–6861, 2015.
41. PANTANO, D. A.; KLEIN, M. L. Characterization of membrane-protein interactions for the leucine transporter from *Aquifex aeolicus* by molecular dynamics calculations. **The Journal of Physical Chemistry B**, v. 113, n. 42, p. 13715–13722, 2009.
42. PETRACHE, H. I.; DODD, S. W.; BROWN, M. F. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by ^2H NMR spectroscopy. **Biophysical Journal**, v. 79, n. 6, p. 3172–3192, 2000.

43. KUCERKA, N.; NIEH, M.-P.; KATSARAS, J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. **Biochimica et Biophysica Acta**, v. 1808, n. 11, p. 2761–2771, 2011.

44. ESSMANN, U.; BERKOWITZ, M. L. Dynamical properties of phospholipid bilayers from computer simulation. **Biophysical Journal**, v. 76, n. 4, p. 2081–2089, 1999.

45. LINDAHL, E.; EDHOLM, O. Molecular dynamics simulation of NMR relaxation rates and slow dynamics in lipid bilayers. **The Journal of Chemical Physics**, v. 115, n. 10, p. 4938–4950, 2001.

46. HOFSTÄÄ, C.; LINDAHL, E.; EDHOLM, O. Molecular dynamics simulations of phospholipid bilayers with cholesterol. **Biophysical Journal**, v. 84, n. 4, p. 2192–2206, 2003.

47. LEACH, A. R. **Molecular modelling: principles and applications**. 2. ed. Harlow, England: Prentice Hall, 2001. 744 p.

48. MARTÍNEZ, L.; BORIN, I. A.; SKAF, M. S. Fundamentos de simulação por dinâmica molecular. In: MORGON, N. H.; COUTINHO, K. (Ed.). **Métodos de química teórica e modelagem molecular**. São Paulo: Editora Livraria da Física, 2007. p. 413–452.

49. CASE, D. A.; BERRYMAN, J. T.; BETZ, R. M.; CERUTTI, D. S.; CHEATHAM III, T. E.; DARDEN, T. A.; DUKE, R. E.; GIESE, T. J.; GOHLKE, H.; GOETZ, A. W.; HOMEYER, N.; IZADI, S.; JANOWSKI, P.; KAUS, J.; KOVALENKO, A.; LEE, T. S.; LEGRAND, S.; LI, P.; LUCHKO, T.; LUO, R.; MADEJ, B.; MERZ, K. M.; MONARD, G.; NEEDHAM, P.; NGUYEN, H.; NGUYEN, H. T.; OMELIAN, I.; ONUFRIEV, A.; ROE, D. R.; ROITBERG, A.; SALOMON-FERRER, R.; SIMMERLING, C. L.; SMITH, W.; SWAILS, J.; WALKER, R. C.; WANG, J.; WOLF, R. M.; WU, X.; YORK, D. M.; KOLLMAN, P. A. **AMBER 2015**. Versão 2015. [S.I.]: University of California, San Francisco, 2015. Programa de computador.

50. BROOKS, B. R.; BROOKS III, C. L.; MACKERELL JR, A. D.; NILSSON, L.; PETRELLA, R. J.; ROUX, B.; WON, Y.; ARCHONTIS, G.; BARTELS, C.; BORESCH, S.; CAFLISCH, A.; CAVES, L.; CUI, Q.; DINNER, A. R.; FEIG, M.; FISCHER, S.; GAO, J.; HODOSCEK, M.; IM, W.; KUCZERA, K.; LAZARIDIS, T.; MA, J.; OVCHINNIKOV, V.; PACI, E.; PASTOR, R. W.; POST, C. B.; PU, J. Z.; SCHAEFER, M.; TIDOR, B.; VENABLE, R. M.; WOODCOCK, H. L.; WU, X.; YANG, W.; YORK, D. M.; KARPLUS, M. CHARMM: the biomolecular simulation program. **Journal of Computational Chemistry**, v. 30, n. 10, p. 1545–1614, 2009.

51. VAN DER SPOEL, D.; LINDAHL, E.; HESS, B.; GROENHOF, G.; MARK, A. E.; BERENDSEN, H. J. GROMACS: fast, flexible, and free. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1701–1718, 2005.
52. PHILLIPS, J. C.; BRAUN, R.; WANG, W.; GUMBART, J.; TAJKHORSHID, E.; VILLA, E.; CHIPOT, C.; SKEEL, R. D.; KALÉ, L.; SCHULTEN, K. Scalable molecular dynamics with NAMD. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1781–1802, 2005.
53. MARRINK, S. J.; TIELEMAN, D. P. Perspective on the Martini model. **Chemical Society Reviews**, v. 42, n. 16, p. 6801–6822, 2013.
54. KLAUDA, J. B.; VENABLE, R. M.; FREITES, J. A.; O'CONNOR, J. W.; TOBIAS, D. J.; MONDRAGON-RAMIREZ, C.; VOROBYOV, I.; MACKERELL JR, A. D.; PASTOR, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. **The Journal of Physical Chemistry B**, v. 114, n.23, p. 7830–7843, 2010.
55. KLAUDA, J. B.; BROOKS, B. R.; MACKERELL JR, A. D.; VENABLE, R. M.; PASTOR, R. W. An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer. **The Journal of Physical Chemistry B**, v. 109, n. 11, p. 5300–5311, 2005.
56. WANG, J.; WOLF, R. M.; CALDWELL, J. W.; KOLLMAN, P. A.; CASE, D. A. Development and testing of a general amber force field. **Journal of Computational Chemistry**, v. 25, n. 9, p. 1157–1174, 2004.
57. TESSIER, M. B.; DEMARCO, M. L.; YONGYE, A. B.; WOODS, R. J. Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. **Molecular Simulation**, v. 34, n. 4, p. 349–363, 2008.
58. PARK, S.; BEAVEN, A. H.; KLAUDA, J. B.; IM, W. How tolerant are membrane simulations with mismatch in area per lipid between leaflets? **Journal of Chemical Theory and Computation**, v. 11, n. 7, p. 3466–3477, 2015.
59. LEE, J.; CHENG, X.; SWAILS, J. M.; YEOM, M. S.; EASTMAN, P. K.; LEMKUL, J. A.; WEI, S.; BUCKNER, J.; JEONG, J. C.; QI, Y.; JO, S.; PANDE, V. S.; CASE, D. A.; BROOKS III, C. L.; MACKERELL JR, A. D.; KLAUDA, J. B.; IM, W. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. **Journal of Chemical Theory and Computation**, v. 12, n. 1, p. 405–413, 2016.

60. JO, S.; KIM, T.; IYER, V. G.; IM, W. CHARMM-GUI : a web-based graphical user interface for CHARMM. **Journal of Computational Chemistry**, v. 29, n. 11, p. 1859–1865, 2008.
61. WU, E. L.; CHENG, X.; JO, S.; RUI, H.; SONG, K. C.; DÁVILA-CONTRERAS, E. M.; QI, Y.; LEE, J.; MONJE-GALVAN, V.; VENABLE, R. M.; KLAUDA, J. B.; IM, W. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. **Journal of Computational Chemistry**, v. 35, n. 27, p. 1997–2004, 2014.
62. JO, S.; KIM, T.; IM, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. **PLoS ONE**, v. 2, n. 9, p. e880, 2007.
63. HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, 1996.
64. FILIPPOV, A.; ORADD, G.; LINDBLOM, G. Influence of cholesterol and water content on phospholipid lateral diffusion in bilayers. **Langmuir**, v. 19, n. 16, p. 6397–6400, 2003.
65. SEELIG, J.; WAESPE-SARCEVIC, N. Molecular order in cis and trans unsaturated phospholipid bilayers. **Biochemistry**, v. 17, n. 16, p. 3310–3315, 1978.
66. PERLY, B.; SMITH, I. C. P.; JARRELL, H. C. Effects of the replacement of a double bond by a cyclopropane ring in phosphatidylethanolamines: a deuterium NMR study of phase transitions and molecular organization. **Biochemistry**, v. 24, n. 4, p. 1055–1063, 1985.
67. FERREIRA, T. M.; CORETA-GOMES, F.; OLLILA, O. H.; MORENO, M. J.; VAZ, W. L.; TOPGAARD, D. Cholesterol and POPC segmental order parameters in lipid membranes: solid state ^1H - ^{13}C NMR and MD simulation studies. **Physical Chemistry Chemical Physics**, v. 15, n. 6, p. 1976–1989, 2013.

APÊNDICE

APÊNDICE A – Valores experimentais de APL para bicamada lipídica composta por POPC.⁴³

APÊNDICE B – Valores experimentais de EB para bicamada lipídica composta por POPC.⁴³

APÊNDICE C – Valores experimentais de DL para bicamada lipídica composta por POPC.⁶⁴

APÊNDICE D – DQM para as simulações por DM dos sistemas na temperatura de 310 K.

POPC

DOPE

POPS

CONCLUSÃO GERAL

Os modelos QSAR propostos neste trabalho se mostraram promissores para a predição da atividade biológica de antagonistas do Sig-1R, descrevendo tanto relações lineares quanto não lineares entre um conjunto de descritores e a atividade biológica dos compostos. Além de propor modelos de QSAR com boa capacidade preditiva, os estudos descritos no Capítulo 1 foram importantes para o aperfeiçoamento da metodologia empregada, que está sendo utilizada em outros trabalhos desenvolvidos no Grupo de Química Quântica do Instituto de Química de São Carlos, USP. Neste trabalho foi proposta uma metodologia para seleção de variáveis e consequente utilização das técnicas de PLS e ANN que se mostrou promissora também para os estudos com outros grupos de moléculas que apresentam atividades biológicas para diferentes receptores.

Em relação aos estudos desenvolvidos no Capítulo 2, foram propostas bicamadas com composição lipídica semelhante às membranas sinápticas, além de terem sido investigados os efeitos da presença de moléculas de colesterol nessas bicamadas. As bicamadas lipídicas obtidas já estão sendo utilizadas em outros estudos, agora envolvendo receptores como o receptor muscarínico e moléculas de neurotransmissores. Além disso, a metodologia aqui empregada também foi utilizada para a construção de outras bicamadas lipídicas cuja composição se assemelha à composição experimental de membranas mitocondriais. Portanto, o trabalho desenvolvido em muito contribuiu para o avanço nas pesquisas de sistemas membranares e em breve será disponibilizado para acesso público.

A formação de pesquisadores na área de Química Medicinal Computacional precisa abranger a utilização de diferentes abordagens para a solução dos problemas investigados. Muitas vezes é necessário utilizar técnicas combinadas para que se possa responder às perguntas sobre algum sistema biológico de interesse. Nesse propósito, o presente trabalho possibilitou que diferentes metodologias fossem aplicadas (aperfeiçoadas), e que, juntamente com os conhecimentos prévios de outras abordagens, transformam-se em uma estratégia poderosa para o avanço das pesquisas em Química Medicinal.