



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE QUÍMICA
CURSO DE LICENCIATURA EM QUÍMICA

ESTERFANIA LAÍS DA SILVA

**OBTENÇÃO DE MODELOS QSAR NA PLATAFORMA LIVRE DE
COMPUTAÇÃO EM NUVEM OCHEM E UMA APLICAÇÃO PARA
ATIVIDADE ACARICIDA DE MONOTERPENOS**

Recife
2022

ESTERFANIA LAÍS DA SILVA

**OBTENÇÃO DE MODELOS QSAR NA PLATAFORMA LIVRE DE
COMPUTAÇÃO EM NUVEM OCHEM E UMA APLICAÇÃO PARA
ATIVIDADE ACARICIDA DE MONOTERPENOS**

Monografia apresentada a coordenação do curso de Licenciatura em Química da Universidade Federal Rural de Pernambuco, como parte dos requisitos para obtenção do título de Licenciada em Química.

Discente: Esterfania Laís da Silva
Orientador: Prof. Dr. Hélcio José Batista

Recife

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- S586o Silva, Esterfania Laís
Obtenção de modelos QSAR na plataforma livre de computação em nuvem OCHEM e uma aplicação para atividade acaricida de monoterpenos / Esterfania Laís Silva. - 2022.
60 f. : il.
- Orientador: Helcio Jose Batista.
Inclui referências, apêndice(s) e anexo(s).
- Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Licenciatura em Química, Recife, 2022.
1. QSAR. 2. OCHEM. 3. Terpenos. 4. Inseticidas botânicos. 5. Contato residual. I. Batista, Helcio Jose, orient. II. Título

ESTERFANIA LAÍS DA SILVA

**OBTENÇÃO DE MODELOS QSAR NA PLATAFORMA LIVRE DE
COMPUTAÇÃO EM NUVEM OCHEM E UMA APLICAÇÃO PARA
ATIVIDADE ACARICIDA DE MONOTERPENOS**

Aprovado em: 08 de junho de 2022.

Banca Examinadora

Prof. Dr. Hécio José Batista

UFRPE - Universidade Federal Rural de Pernambuco

Prof. Dr. Marcilio Martins de Moraes

UFRPE - Universidade Federal Rural de Pernambuco

Prof. Dr. José Euzebio Simões Neto

UFRPE - Universidade Federal Rural de Pernambuco

A toda minha família, amigos e as pessoas que me incentivaram e ajudaram em todas as etapas deste processo.

AGRADECIMENTOS

Creio que primeiramente devo agradecer a Deus, pois sem ele não teria o dom da vida e não estaria passando por essa etapa importante, pois não foi fácil, mas sei que sempre estive comigo durante todo tempo. Também agradeço a minha família (avó, tia, mãe, irmãos, tios, primas, etc.), todos, independente da forma ou modo, sempre me ajudaram e me incentivaram durante todo esse processo.

A mesmo tempo também quero agradecer aos meus amigos, todos em geral que sempre estiveram comigo durante esse crescimento acadêmico. A toda família LQ3 (filhos de Lavoisier), que tenho o maior orgulho de cada um e o prazer de conhecê-los e tê-los na minha vida.

Durante toda caminhada acadêmica, encontrei diversos tipos de pessoas e muitas delas foram e são importantes para mim. Mas em especial, queria agradecer aos meus amigos Reobe, que durante todos esses anos de graduação e na vida, sempre me ajudou de todas as formas, a Fernando, que independentemente do tempo ou espaço sempre esteve e está comigo, nos meus piores e melhores momentos da vida acadêmica e pessoal. As minhas amigas Nayara, que me iluminou deste o início, Daniele, que me ensinou a nunca desistir, mesmo que fique difícil, e Thiara, que sempre teve e tem um conselho para mim.

E por fim, e não menos importante a todos os meus professores, do ensino regular ao acadêmico, pois com cada um aprendi e construí uma parte de mim. Ao meu professor e orientador Dr. Hércio José, que aceitou me guiar e orientar durante esta etapa acadêmica, só tenho que agradecer.

Durante todos esses anos, passei por diversas coisas e muitas delas quase me fizeram desistir, mas no final consegui chegar, por isso, agradeço a todos que me ajudaram de qualquer forma, obrigado por tudo.

LISTA DE FIGURAS

Figura 1	Processo de obtenção e processamento de descritores moleculares e várias aplicações	28
Figura 2	Imagem da plataforma web OCHEM na página inicial	38
Figura 3	Imagem da plataforma web OCHEM na página inicial	38
Figura 4	Fazendo o upload do arquivo com os dados da série de treino	39
Figura 5	Pré-seleção de opções padrão da OCHEM (recomendadas)	39
Figura 6	Escolha dos descritores, no caso os da versão 7 do DRAGON	40
Figura 7	Imagem após o procedimento dos cálculos	40
Figura 8	Seleção dos dados e escolha do formato do documento a ser baixado	41
Figura 9	Adicionando o arquivo com as estruturas da série de treino	42
Figura 10	Adição e envio do documento na plataforma web do OCHEM	42
Figura 11	Escolhendo no menu “Models” a opção “criar um modelo”	43
Figura 12	Adição/validação dos descritores e escolha dos métodos para a construção do modelo	43
Figura 13	Continuação da escolha dos métodos para a construção do modelo e escolha do modo de avaliação	44
Figura 14	Seleção do documento contendo os dados	44
Figura 15	Seleção do pré-procedimento dos dados	44
Figura 16	Seleção dos descritores para construção do modelo	45
Figura 17	Continuação da seleção dos descritores para construção do modelo	45
Figura 18	Seleção para filtragem dos descritores	46
Figura 19	Configuração para o modelo	46
Figura 20	Início do cálculo para o modelo	46
Figura 21	Energia e forma do orbitais HOMO e LUMO do composto α -pineno	47
Figura 22	Energia e forma do orbitais HOMO e LUMO do composto p -cimeno	48
Figura 23	Energia e forma do orbitais HOMO e LUMO do composto β -pineno	48
Figura 24	Energia e forma do orbitais HOMO e LUMO do composto α -terpineol	49
Figura 25	Energia e forma do orbitais HOMO e LUMO do composto geraniol	49

Figura 26	Energia e forma do orbitais HOMO e LUMO do composto R-limoneno	50
Figura 27	Energia e forma do orbitais HOMO e LUMO do composto 1-8-cineole	50
Figura 28	Energia e forma do orbitais HOMO e LUMO do composto R-linalol	51
Figura 29	Energia e forma do orbitais HOMO e LUMO do composto 4-(S)terpineol	51
Figura 30	Energia e forma do orbitais HOMO e LUMO do composto mentol	52
Figura 31	Energia e forma do orbitais HOMO e LUMO do composto acetato de terpenila	52
Figura 32	Energia e forma do orbitais HOMO e LUMO do composto terpinoleno	53
Figura 33	Estruturas no formato SMILES de nove compostos usados na série de treino	54
Figura 34	Valores dos descritores para o Dragon versão 7 para o método AM1	55
Figura 35	Valores dos descritores para o Dragon versão 7 para o método B3LYP	55
Figura 36	Quadro resumo obtido da plataforma OCHEM	57

LISTA DE ABREVIATURAS E SIGLAS

EQS	Equação de Schrödinger
NIST	National Institute of Standards and Technology
OCHEM	Online Chemical Modeling Environment
DFT	Teoria Funcional da Densidade (<i>Density Functional Theory</i>)
HFR	Método Hartree-Fock-Roothaan
IUPAC	União Internacional da Química Pura e Aplicada
MSE	Método Semi-Empírico
PCM	PCM (<i>Polarizable Continuum Model</i>)
QSAR	Relação Quantitativa Estrutura-Atividade (<i>Quantitative Structure-Activity Relationship</i>)
MLR	Regressão Linear Múltipla (<i>multiple linear regression</i>)
PCA	Análise de componentes principais (<i>Principal component analysis</i>)
PLS	Mínimos Quadráticos Parciais (<i>partial least squares</i>)
ANN	Redes Neurais (<i>neural networks</i>)
PLS	Mínimos Quadráticos Parciais (<i>partial least squares</i>)
LAQTEO	Laboratório de Química Computacional do Departamento de Química da UFRPE
LPNBio	Laboratório de Produtos Naturais e Bioativos
UFRPE	Universidade Federal Rural de Pernambuco

LISTA DE EQUAÇÕES

Equação 1	Equação de Schrödinger Independente do tempo	23
Equação 2	Operador Hamiltoniano Completo	24
Equação 3	Equação de Schrödinger Nuclear	24
Equação 4	Equação de Schrödinger Eletrônica	24
Equação 5	Função de onda expressa como um determinante de Slater	25
Equação 6	Modelo Ideal de Regressão	31

A química é necessariamente uma ciência
experimental: as conclusões são
extraídas de dados e seus princípios são
apoiados pela evidência dos fatos.
(Michael Faraday)

RESUMO

No presente trabalho é apresentado o uso dos recursos da plataforma da web de computação em nuvem OCHEM, do inglês *Online Chemical Modeling Environment*, de acesso livre e desenvolvida para armazenamento de dados, desenvolvimento de modelos e publicação de informação Química. Como exemplo de aplicação na plataforma, procuramos obter modelos QSAR – Relação Quantitativa Estrutura-Atividade para atividade inseticida sobre a espécie do Ácaro Rajado (*Tetranychus Urticae*), uma praga comum nas lavouras de agricultura familiar na Região Nordeste do Brasil, de uma série de compostos da classe de monoterpenos: α -pineno, p -cimeno, β -pineno, mentol, α -terpineol, geraniol, R-limoneno, 1-8-cineole, R-linalol, acetado de terpenila, e terpinoleno e 4-(S) ter-pineol. Para este fim, realizamos uma revisão bibliográfica do uso de monoterpenos no controle de pragas na agricultura, bem como uma revisão do método Relação, QSAR, do inglês *Quantitative Structure-Activity Relation-ship*. Para a aplicação da plataforma OCHEM na obtenção de modelos QSAR para a série de monoterpenos realizamos a modelagem molecular por métodos de Química Quântica da série, gerando as estruturas moleculares com o objetivo de alimentar a plataforma OCHEM, para cálculo dos descritores moleculares e obtenção de modelos QSAR. Fizemos a exploração de diferentes conjuntos de descritores bem como diferentes tipos de métodos de aprendizagem de máquina, como Regressão Linear Múltipla (MLR – *Multiple Linear Regression*), Mínimos Quadrados Parciais (PLS – *Partial Least Squares*) e Redes neurais (NN - *Neural Networks*). Os modelos obtidos para a série em estudo, no entanto, não tiveram qualidade estatística suficiente devido provavelmente, primeiro à quantidade muito pequena de compostos na série de treino do modelo com dados experimentais disponíveis e, também, ao provável grau de não linearidade presente na relação entre os descritores e a atividade.

Palavras-chave: QSAR. OCHEM. Terpenos. Inseticidas botânicos. Contato residual.

ABSTRACT

The present work focuses in using the resources of the free access cloud computing web platform OCHEM - English Online Chemical Modeling Environment, designed for data storage, development of models and publication of chemical information. As an example of application on the platform, we sought to obtain QSAR models - Quantitative Structure-Activity Relation for insecticidal activity on the species of Spider Mite (*Tetranychus Urticae*), a common pest in family farming in the Northeast region of Brazil, from a series of compounds of the monoterpene class: α -pinene, p-cymene, β -pinene, menthol, α -terpineol, geraniol, R-limonene, 1-8-cineole, R-linalool, terpenyl acetate, and terpinolene and 4- (S) ter-pineol. To this end, we carried out a literature review on the monoterpenes usage in agriculture pest control, as well as a review of the Quantitative Structure-Activity Relation-ship method, QSAR. For the application of the OCHEM platform to obtain QSAR models for the series of monoterpenes, we performed the molecular modeling of the series compounds by methods of Quantum Chemistry, generating the molecular structures in order to feed the OCHEM platform, to calculate the molecular descriptors and obtain of QSAR models. We explored different sets of descriptors as well as different types of machine learning methods, such as Multiple Linear Regression (MLR), Partial Least Squares (PLS) and Neural Networks (NN - Neural Networks). The models so obtained for the series under study, however, did not have sufficient statistical quality, probably due, firstly, to the small number of compounds in the training set with available experimental data and, also, to a likely strong degree of non-linearity present in the relationship between the chosen descriptors and the activity.

Keywords: QSAR. OCHEM. terpenes. Botanical insecticides. Residual contact.

SUMÁRIO

1 INTRODUÇÃO	15
2 FUNDAMENTAÇÃO TEÓRICA	19
2.1 Monoterpenos	19
2.2 Modelagem Molecular por Métodos de Química Quântica	21
2.3 Química Computacional	26
2.4 Quimiometria e Modelos QSAR	26
2.5 Modelos QSAR Baseados em Regressão Linear Múltipla	30
2.6 Modelos QSAR Baseados em outros Métodos de <i>Machine Learning</i>	31
3 METODOLOGIA	33
3.1 Obtenção Dados Experimentais de Contato Residual	33
3.2 Procedimento Teórico	34
3.3 Construção e Validação dos Modelos QSAR	40
4 ESTRUTURA MOLECULAR E ELETRÔNICA DOS COMPOSTOS	46
5 APLICAÇÃO DA OCHEM PARA OBTENÇÃO DE MODELOS QSAR PARA UMA SÉRIE DE MONOTERPENOS	53
5.1 Ilustração do arquivo de saída da OCHEM com os descritores Físico- Químicos	53
5.2 Ilustração do arquivo de saída da OCHEM com alguns modelos QSAR	55
6 CONCLUSÃO E PERSPECTIVAS	57
REFERÊNCIAS	58
APÊNDICE	

1 INTRODUÇÃO

Existe atualmente uma urgência em relação a resultados práticos da Ciência, acelerando descobertas pelo uso do método científico ao produzir uma abordagem mais ágil e rigorosa para enfrentar problemas complexos, buscando soluções para os desafios mais urgentes do mundo, como as mudanças climáticas, novas doenças, dentre outros temas (IBM RESEARCH, 2021). Acelerar e dimensionar a descoberta hoje é mais importante do que nunca, e as transformações na Ciência e nos outros segmentos da sociedade exigem o desenvolvimento de tecnologia da informação em novos modelos que atendam a essa necessidade. Uma tecnologia que se destaca neste sentido é a computação em nuvem (IBM RESEARCH, 2021).

O processo de transferência de dados de *data centers* para desktops para análise técnica ou científica, ainda bastante comum, logo se tornará impraticável pelo crescimento acelerado do volume de conjuntos de dados científicos. Em vez disso, o processamento passará gradativamente a ocorrer num novo modelo caracterizado por servidores de alto desempenho com dados distribuídos, como a computação em nuvem, que é uma nova maneira de adquirir recursos de computação e armazenamento sob demanda por meio de tecnologias de virtualização. Há relatos de bons resultados da aplicabilidade da computação em nuvem à computação científica, com a possibilidade de escolha de quais tipos de aplicativos podem ser executados de forma barata e eficiente na nuvem. Um exemplo de importante aplicação é o processamento de um grande conjunto de dados para criar um novo produto científico (BERRIMAN, 2013).

De acordo com o NIST - *National Institute of Standards and Technology*, a computação em nuvem é um modelo para permitir, convenientemente, o acesso à rede de internet, sob demanda, a um conjunto compartilhado de recursos computacionais configuráveis (por exemplo, redes, servidores, armazenamento, aplicativos e serviços) que podem ser rapidamente provisionados e liberados com mínimo esforço de gestão ou interação com o prestador de serviços (NIST, 2011).

Uma das linhas de pesquisa em Química que se beneficia dos recursos da computação em nuvem, devido ao seu fácil acesso, baixo custo e rapidez, é o

planejamento racional de compostos com alguma atividade biológica de interesse, particularmente fármacos ou produtos de uso agrícola. O termo “planejamento racional” refere-se à busca por novos compostos por meio do uso de ferramentas de modelagem computacional de potenciais propriedades a partir de um grande número de dados empíricos ou teóricos de compostos já conhecidos, tratados por métodos estatísticos altamente sofisticados (análise multivariada). Métodos já bastante tradicionais para este fim são, por exemplo, a regressão linear múltipla (MLR, do inglês *multiple linear regression*), análise de componentes principais (PCA, do inglês *principal component analysis*), mínimos quadrados parciais (PLS, do inglês *partial least squares*), dentre outros, os quais são de emprego frequente na área chamada Quimiometria (BARROS NETO, 1995). Mais modernamente, tem ficado bastante conhecido o termo “aprendizagem de máquina”, do inglês *machine learning*, que constitui uma ferramenta de programação computacional cada vez mais eficiente para o desenvolvimento de dispositivos de interatividade homem-máquina bem como de métodos que possam prever as propriedades de classes de novos compostos propostos, antes mesmo que estes sejam sintetizados, gerando grande economia de recursos e de tempo (WANG, 2021). A aprendizagem de máquina geralmente envolve construir um modelo, treinar o modelo, realizar a validação, repetir o treinamento e a validação até que um modelo adequado seja obtido e, finalmente, testar o modelo em dados não expostos anteriormente ao modelo. O objetivo de um modelo de aprendizagem de máquina é selecionar padrões dos dados de entrada ou generalizar esses dados e aplicar os resultados a dados de teste desconhecidos (WANG, 2021). As redes neurais artificiais (ANN, do inglês *artificial neural networks*), ou simplesmente redes neurais, são um importante método que se destaca na linha de aprendizado de máquina e têm sido empregadas também na Química (WANG, 2021).

As ferramentas quimiométricas mencionados acima, MLR, PCA, PLS e NN, encontram uma importante aplicação nos métodos chamados Relação Quantitativa Estrutura-Propriedade (QSAR, do inglês *Quantitative Structure-Activity Relationship*), usados para descrever quantitativamente as relações entre a estrutura Química de moléculas e sua atividade biológica o mesmo, permitindo o planejamento racional ou o reposicionamento de fármacos. Para isso é necessário um número total de variáveis, que pode ser muito maior que o número que será efetivamente incluído nos modelos (MACHADO, 2022; TAVARES, 2004). Aliando esta linha de aplicação em QSAR aos

recursos de computação em nuvem, destacamos a plataforma *web* de computação em nuvem OCHEM (SUSHKO, 2011), de acesso livre e desenvolvida para armazenamento de dados, desenvolvimento de modelos e publicação de informação Química, a qual será o foco principal deste trabalho. Neste sentido, tem-se como **objetivo principal**:

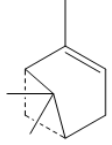
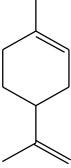
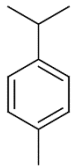
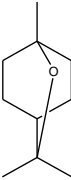
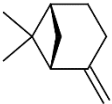
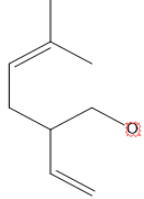
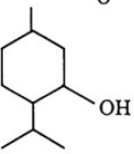
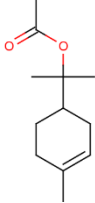
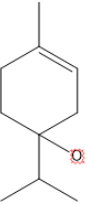
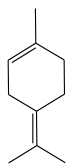
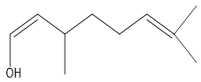
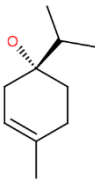
Demonstrar o uso da plataforma OCHEM na obtenção de modelos QSAR para uma aplicação particular, qual seja a atividade acaricida sobre a espécie do ácaro rajado (*Tetranychus urticae*), uma praga comum nas lavouras de agricultura familiar na Região Nordeste do Brasil, de uma série de compostos da classe de monoterpenos (da CAMARA, 2020).

Como **objetivos específicos** tem-se:

1. Realizar a aplicação da plataforma OCHEM para obtenção de modelos QSAR para a atividade acaricida sobre o *Tetranychus Urticae* (ácaro rajado) da seguinte série de monoterpenos: α -pineno, *p*-cimeno, β -pineno, mentol, α -terpineol, geraniol, R-limoneno, 1-8-cineole, R-linalol, acetado de terpenila, terpinoleno, 4-(S)-terpineol.
2. Realizar a modelagem molecular por métodos de Química quântica da série de monoterpenos, gerando as estruturas moleculares com o objetivo de alimentar a plataforma OCHEM, para cálculo dos descritores moleculares e obtenção de modelos QSAR. Calcular diferentes conjuntos de descritores moleculares para a série de terpenos em estudo, numa investigação daqueles conjuntos mais adequados para a obtenção de modelos QSAR com boa capacidade preditiva.
3. Obter modelos QSAR usando a plataforma OCHEM para a atividade acaricida da série de monoterpenos em estudo, por meio de diferentes tipos de métodos de aprendizagem de máquina (*machine learning*).

Os dados experimentais de atividade acaricida para os monoterpenos constam na tabela 1 e foram cedidos ao nosso laboratório, o LAQTEO/UFRPE (Laboratório de Química Teórica - UFRPE), coordenado pelo orientador deste trabalho, pelo grupo de pesquisa dos professores Dr. Cláudio Câmara e Dr. Marcílio Martins de Moraes, do Laboratório de Produtos Naturais Bioativos - UFRPE).

Tabela 1. Série de monoterpenos estudada e sua atividade acaricida por contato residual.

(Nº)/nome composto	Estrutura	Atividade Contato Residual (intervalo de confiança estatística)	(Nº)/nome composto	Estrutura	Atividade Contato Residual (intervalo de confiança estatística)
(1) α -pineno		34,13 (26,28-44,81)	(7) R-Limoneno		76,94 (63,19-92,20)
(2) <i>p</i> -cimeno		36,95 (28,03-46,49)	(8) 1,8-Cineole		91,80 (72,78-117,84)
(3) β -pineno		38,93 (31,72-46,91)	(9) R-linalol		138,88 (78,57-200,22)
(4) Mentol		39,56 (31,12-48,89)	(10) Acetato de terpenila		151,54 (129,54-179,85)
(5) α -terpineol		40,43 (27,48-58,72)	(11) Terpinoleno		341,90 (206,90-520,83)
(6) Geraniol		47,15 (32,72-64,15)	(12) 4-(S) terpineol		366,56 (306,47-421,30)

Fonte: autora

2 FUNDAMENTAÇÃO TEÓRICA

2.1 MONOTERPENOS

O uso de produtos extraídos de plantas para o controle de pragas na agricultura é bastante antigo na humanidade. Por anos, o Brasil foi um grande exportador de produtos naturais usados como inseticidas para lavouras tais como piretro, nicotina e rotenona. Mas esses inseticidas apresentavam problemas de baixa eficiência, o que ocasionou o advento dos inseticidas sintéticos, substituindo por sua vez, os inseticidas naturais no tratamento contra pragas e insetos na agricultura (CORREA, 2011).

Desde meados do século XX, os inseticidas são utilizados para matar, atrair ou repelir insetos. A descoberta dessas substâncias, o isolamento, a síntese, a avaliação toxicológica e o impacto ambiental dos inseticidas constituem um vasto tópico de pesquisas no mundo inteiro. No controle químico de pragas, em particular de insetos, a mera toxicidade de uma substância não necessariamente a qualifica como um inseticida. Para uma substância ser considerada um inseticida, deve-se levar em conta diversos fatores como alta eficiência em pequenas concentrações, ausência de toxicidade e bioacumulação em tecidos animais, fácil obtenção, além da viabilidade econômica (VIEGAS, 2003).

No entanto, tem havido um crescimento em danosos impactos ambientais a partir do uso dos produtos sintéticos de alta eficiência inseticida, os quais apresentam elevados tempos de permanência no solo, contaminação de aquíferos, e bioacumulação em animais, com severas consequências para a saúde da população, o que tem forçado os governos e órgãos de pesquisa científica a investirem estratégias alternativas que sejam ambientalmente sustentáveis (CORREA, 2011). Portanto, o retorno ao uso de produtos naturais ou modificados sinteticamente que possuam atividade inseticida (inseticidas botânicos), de forma combinada com a utilização de outras práticas agrárias sustentáveis, tem merecido bastante atenção nos últimos tempos.

É um fato conhecido que as plantas possuem enorme capacidade de adaptação ao ambiente e, em um processo de seleção natural, conseguem tornar-se resistentes a diversas situações sob ação de agentes agressores. Um exemplo é sua

própria proteção contra pragas e insetos proveniente de várias substâncias bioativas presentes nelas, como por exemplo os compostos da classe dos monoterpenos, comumente encontrados nas plantas, que apresentam atividade inseticida (VIEGAS, 2003).

Devido à descoberta empírica, desde a antiguidade, dessas propriedades, extratos de tais vegetais têm sido utilizados como importantes agentes no controle de pragas em culturas agrícolas (BRAIBANTE, 2012). Esta prática contribuiu para o desenvolvimento da Química e deu origem a uma área de pesquisa científica chamada Química de Produtos Naturais (BERLINCK, 2017). Compostos com atividade biológica inseticida, tais como muitos da classe dos terpenos, são chamados de inseticidas naturais.

Na literatura é descrito que os terpenoides possuem atividade em insetos como inibidores ou retardadores de crescimento, causam danos na maturação, redução da capacidade reprodutiva, podendo levar os insetos predadores à morte por inanição ou toxicidade direta. A ação inseticida destes terpenoides é decorrente da inibição da acetilcolinesterase, enzima importante no metabolismo dos insetos. Vários monoterpenos foram isolados e avaliados quanto à toxicidade frente a diferentes insetos. Estes estudos envolveram α -pineno, β -pineno, 3-careno, limoneno, mirceno, α -terpineno e canfeno (VIEGAS, 2003).

Para Arroio (2010) as interações de um composto com seu receptor biológico são determinadas por forças intermoleculares, ou seja, interações lipofílicas, polares, eletrostáticas e estéricas. Sendo assim, substâncias que apresentam propriedades de estrutura/atividade, ao interagirem com um alvo específico, devem possuir uma estrutura de forma que a disposição de seus grupos funcionais favoreça uma complementaridade ao sítio de ligação. Ou seja, quando maior a absorção, maior será a afinidade e sua atividade biológica. É neste contexto que surge a importância dos métodos QSAR, que buscam modelar a relação entre a estrutura e a atividade, descrevendo esse tipo de interação entre um ligante e o seu receptor biológico por meio de um conjunto de propriedades moleculares (descritores moleculares), uma vez que estas propriedades estão diretamente relacionadas às forças intermoleculares envolvidas na interação ligante-receptor. Os descritores moleculares representam

uma enorme ferramenta para prever as propriedades das substâncias, classificar suas estruturas químicas ou procurar similaridade entre elas.

2.1 MODELAGEM MOLECULAR POR MÉTODOS DE QUÍMICA QUÂNTICA

2.2.1 Mecânica Quântica

O pensamento de Albert Einstein sobre o modelo para a luz baseado na ideia de fóton e a proposta de Louis de Broglie sobre ondas de matéria, trouxeram novos conceitos, antes desconhecidos e agora imaginados. Mas essas ideias ainda não estavam completas, faltava algo que as complementasse, ou seja, algo que juntasse a velha teoria quântica e a nova mecânica quântica.

A Mecânica Quântica surgiu no século XX como resultado dos estudos de diversos cientistas, mas foi a partir dos resultados obtidos nas suas pesquisas sobre o trabalho de Broglie, que Erwin Schrödinger desenvolveu uma equação (EQS – Equação de Schrödinger) que permitiu que se compreende as propriedades fundamentais das partículas-ondas.

A palavra quântica refere-se ao fato de que, em sistemas de escala de tamanho atômico ou molecular, a energia dos sistemas não varia de forma contínua como a de um objeto macroscópico, e sim de forma discreta, ou seja, de forma quântica ou quantizada (BATISTA, 2020).

Um aspecto fundamental é que não é possível descrever as trajetórias de partículas subatômicas, como núcleos e elétrons, de forma determinada como para corpos macroscópicos, como por exemplo órbitas circulares ou elípticas de um planeta em torno de uma estrela. Ao contrário, só é possível determinar regiões do espaço com a maior *probabilidade* de se encontrar as referidas partículas, ou seja, as *trajetórias mais prováveis*. As *energias* e as regiões de localização mais provável das partículas, chamadas *orbitais*, são resultados da resolução quantitativa da famosa EQS (BATISTA, 2020).

Para essa equação são apresentadas duas versões. A primeira descreve a evolução temporal dos estados do sistema, conhecida como equação de Schrödinger dependente do tempo e a segunda apresenta os estados estacionários do sistema,

cujas energias são quantizadas, apresentada como equação de Schrödinger independentemente do tempo (BELEM, 2016).

A versão independente do tempo da equação se apresenta na forma de uma equação diferencial, apresentada como um problema de autovalor do operador Hamiltoniano do sistema (H), que representa a energia total do sistema. Os estados estacionários, de energias fixas E , são os estados quânticos, e correspondem às autofunções de H, denominadas funções de onda (Ψ) e os correspondentes autovalores de H são as energias E .

Equação 1 - Equação de Schrödinger Independente do Tempo

$$H\Phi(\vec{r}, \vec{R}) = E\Phi(\vec{r}, \vec{R})$$

Fonte: (BELEM,2016).

Na equação acima o \vec{r} representa as coordenadas dos elétrons e \vec{R} a coordenada do núcleo. Contudo ainda não é possível uma resolução analítica desta equação para sistemas multieletrônicos, tornando necessário o uso de aproximações.

2.2.2 Métodos aproximados para resolução da Equação de Schrödinger (EQS)

- Aproximação Born-Oppenheimer (ABO)

Mesmo a equação de Schrödinger representando as propriedades corretas do sistema, ela não apresenta uma resolução analítica para sistemas moleculares (SZABO; OUSTLAND, 1989). No início do século XX os cientistas Max Born e Robert Oppenheimer estudavam a relação núcleo-elétron e chegaram à conclusão que o movimento nuclear, comparado ao do elétron, é lento, adaptando-se a qualquer mudança de posição referente ao núcleo. Sendo assim, o núcleo percebe a presença do elétron a partir de uma nuvem de carga estacionária. Pensando assim, chegaram a conclusão acerca do desacoplamento entre os movimentos nucleares e eletrônicos, possibilitando o uso da equação de Schrödinger para sistemas moleculares e a descrição da estrutura química.

O método de aproximação Born-Oppenheimer é amplamente usado, pois torna bastante prática a resolução da EQS para os elétrons. O operador hamiltoniano (H) inclui os termos de energia cinética dos núcleos e dos elétrons e os potenciais de interação a que estes estão sujeitos e, desta forma, podemos escrever o operador como:

Equação 2 - Operador Hamiltoniano Completo

$$H = T_N + T_e + V_{Ne} + V_{ee} + V_{NN}$$

Fonte: (BELEM,2016).

Em que:

T é o termo de energia cinética;

V é o potencial eletroestático;

N é referente ao núcleo e

“e” representa o termo elétron.

Substituindo o operador hamiltoniano na EQS (eq. 1), chegamos à seguinte equação:

Equação 3 - Equação de Schroedinger Nuclear

$$(T_N + E_e + V_{NN})\Phi_N(\vec{R}) = E_N \Phi_N(\vec{R})$$

Fonte: (BELEM,2016).

Mas vale ressaltar que nesta equação, o potencial para o movimento dos núcleos é a energia eletrônica adicionada a repulsão internuclear. Mas antes, devemos buscar solução para a equação eletrônica.

Equação 4 - Equação de Schrödinger Eletrônica

$$(T_e + V_{Ne} + V_{ee})\Phi_e(\vec{r}; \vec{R}) = E_e(\vec{R})\Phi_e(\vec{r}; \vec{R})$$

Fonte: (BELEM,2016).

A partir daqui, ainda há muitos tópicos a serem estudados e alguns métodos serão descritos posteriormente.

- Método de Hartree-Fock-Roothaan (HFR)

A partir desta equação 4, Douglas Hartree começou a descrever a função de onda $\Phi_e(\vec{r}; \vec{R})$, como um produto de funções de onda mono-eletrônicas, chamadas de orbitais (SZABO; OUSTLAND 1989), tendo somente uma solução exata para sistemas, com apenas um átomo de hidrogênio.

A limitação deste método é de não levar em conta a antissimetria da função de onda. Vladimir Fock e John Slater propuseram a inclusão da antissimetria a partir da introdução de um determinante, chamado de Determinante de Slater (Φ) (ANTUNES, 1999), como descrito na equação a seguir:

Equação 5 - Função de onda expressa como um determinante de Slater

$$\Psi(1, 2, \dots, N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1) & \psi_2(1) & \dots & \psi_N(1) \\ \psi_1(2) & \psi_2(2) & \dots & \psi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(N) & \psi_2(N) & \dots & \psi_N(N) \end{vmatrix}$$

Fonte: (BELEM, 2016).

A partir desta equação temos os estados eletrônicos representados, não considerando o efeito da correlação eletrônica. Utilizando o método de variacional (SZABO; OUSTLAND 1989), podemos chegar às equações de Hartree-Fock (HF), na qual é possível determinar uma estimativa variacional para a energia do estado fundamental do sistema molecular.

A utilização das funções monoelétrônicas, como proposta para solucionar o problema da teoria de HFR, são denominadas de orbitais moleculares, conhecida também como Teoria de Orbitais Moleculares, (baseada no método de Hartree-Fock Roothaan-HFR) (ATKINS, 2012). Baseado na equação de HFR foi desenvolvida a teoria semiempírica de orbitais moleculares (métodos semiempíricos), descritos na seção a seguir (BATISTA, 2020).

O método de HFR não introduz a correção de correlação eletrônica, havendo apenas correlação de spin. Para aplicações gerais é necessário impor aproximações ao método. A correção deste erro é um dos principais desafios da modelagem molecular quântica, como será mostrado a seguir. Uma abordagem para introduzir estas correções de forma simplificada são os métodos semiempíricos.

- Métodos Semiempíricos (MSE)

Criados nos anos de 1930, os métodos semiempíricos tentam descrever a estrutura e propriedades de moléculas a partir da resolução aproximada da Eq. de Schrödinger e com o uso de parâmetros experimentais. Com os métodos semiempíricos, resolve-se a equação de HFR por meio da parametrização das integrais a partir de dados experimentais. Comparados com outros métodos como *ab initio*, o método semiempírico apresenta custo computacional muito menor, porém é mais aproximado, mas sem perder qualidade.

Neste método é aplicada a aproximação ZDO (do inglês *zero differential overlap*), destacando os métodos MNDO, AM1, PM3 (POBLE; BEVERIDGE, 1969). A variedade destes MSE se diferencia mediante o tipo de aplicação, mudando assim, a sua parametrização.

- Teoria do Funcional da Densidade (DFT – *Density Functional Theory*)

Após o método HFR surgir, diversos outros meios para correção de correlação eletrônica foram criados, e essas soluções são chamadas de métodos pós-Hartree-Fock ou métodos correlacionados. Um destes métodos é chamada de Teoria do Funcional da Densidade (MORGON; CUSTÓDIO, 1995), baseia-se na ideia de que a energia total de uma molécula pode ser univocamente representada por um funcional da densidade eletrônica total. Neste caso, uma vez conhecido este funcional, a correlação eletrônica dinâmica estaria automaticamente determinada, e assim todas as propriedades, em princípio, poderiam ser obtidas, (CALVACANTI, 2016). Apesar de não existir um caminho para se determinar tal funcional de forma exata, usam-se vários funcionais desenvolvidos empiricamente, como o B3lyp, um método bastante utilizado e consagrado na literatura após os anos 2000.

O método DFT comparado a outros métodos *ab initio*, apresenta um custo computacional bem mais baixo, porém seus resultados são, em geral, mais confiáveis em relação os métodos semiempíricos, os quais possuem dados parametrizados e certa especificidade para cada propriedade.

2.3 QUÍMICA COMPUTACIONAL

A modelagem molecular consiste na utilização de uma série de métodos teóricos, geralmente métodos de Química quântica, e técnicas computacionais empregados para explicar e prever a estrutura e propriedades moleculares. A modelagem molecular é realizada a partir do uso de vários programas que permitem o cálculo de estruturas e propriedades, visualização e interpretação destas propriedades (RODRIGUES, 2001). Softwares como Gaussian, Hyperchem, etc. são muito utilizados para estes fins e, neste trabalho, serão empregados para a obtenção das estruturas moleculares dos monoterpenos investigados que foram utilizadas para alimentar a plataforma OCHEM, para construção de modelos de QSAR (*Quantitative Structure-Activity Relationship*). Além das estruturas moleculares e eletrônicas, alguns descritores quânticos moleculares são obtidos automaticamente nestes cálculos, no entanto, não foram utilizados na construção dos modelos QSAR.

2.4 QUIMIOMETRIA E MODELOS QSAR

2.4.1 Quimiometria

A Quimiometria é o campo da Química que utiliza técnicas estatísticas e matemáticas para construção e otimização das condições experimentais, e para informações químicas relevantes de dados químicos multivariados. A partir dos métodos quimiométricos podemos transformar propriedades complexas em valores numéricos, por exemplo os descritores moleculares. Todavia, vale ressaltar que esses resultados dependem como o valor numérico é obtido, pois esses métodos podem ter diversas limitações e interpretação diferentes.

2.4.2 Descritores Moleculares

O descritor molecular é o resultado final de um procedimento lógico e matemático que transforma a informação Química codificada dentro de uma representação simbólica de uma molécula em um número útil ou o resultado de algum experimento padronizado. Para este fim são muito utilizados métodos quimiométricos, que buscam sintetizar diferentes propriedades complexas em valores numéricos (descritores moleculares) que codificam uma informação Química (CONSONNI; TODESCHINI, 2010).

É bastante comum a utilização de descritores para desenvolvimento de modelos que relacionam a estrutura com a atividade de uma molécula, ou modelos QSAR, em que é otimizada uma função resposta atrelada à atividade em estudo de uma série de compostos quimicamente similares, usando métodos de otimização, em geral regressão linear múltipla, no qual esta função é expandida linearmente em termos dos descritores relevantes estatisticamente. No quadro 1 tem-se um fluxograma ilustrando o processo geral da obtenção descritores moleculares por meio de diversos métodos, desde a química experimental até métodos matemáticos avançados, seu processamento por métodos estatístico e diversas aplicações.

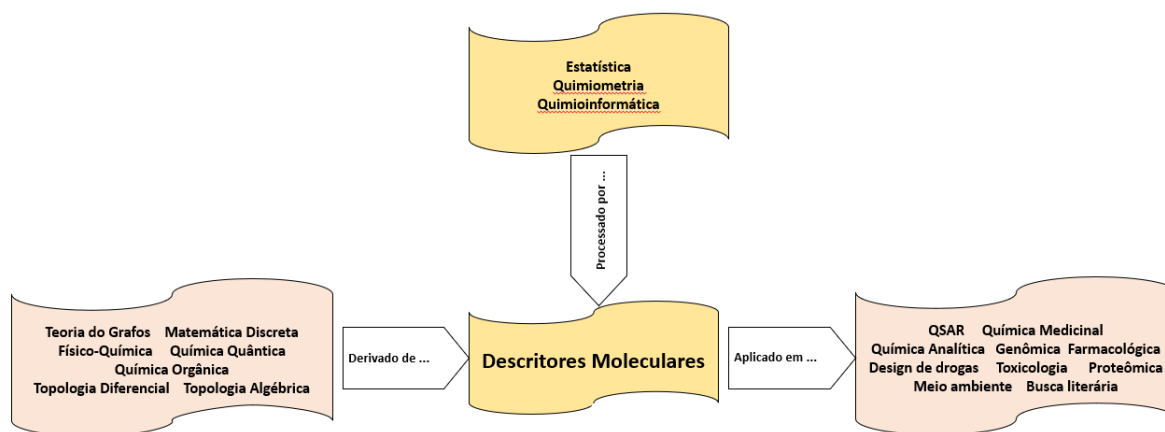


Figura 1. Processo de obtenção e processamento de descritores moleculares e várias aplicações.
Fonte: autora.

Esses descritores moleculares podem ser divididos de acordo com sua natureza (CONSONNI; TODESCHINI 2010) ou como são obtidos.

2.1.1.1 Descritores Físico-Químicos Experimentais

São considerados descritores experimentais aqueles que são obtidos a partir de observação experimental, existindo assim diversos tipos de descritores experimentais, tais como logP, refratividade molar, momento de dipolo, polarizabilidade e diversas propriedades físico-químicas (entalpia, entropia, etc.). É importante ressaltar que estes mesmos descritores exemplificados podem também ser obtidos de maneira teórica.

2.1.1.2 Descritores Teóricos

Os descritores teóricos são aqueles que são obtidos por cálculos teóricos, por diferentes métodos, podendo ser referentes a qualquer propriedade mensurável do composto, aspecto morfológico ou constitucional. Em particular, as propriedades físico-químicas que podem ser utilizadas como descritores experimentais, como dito acima, mas podem também ser estimados por métodos teóricos, em geral por métodos quânticos e termodinâmicos. Alguns exemplos são o coeficiente de partição água/octanol, LogP, que expressa quantitativamente o grau de hidrofobicidade da molécula, o momento de dipolo elétrico, a polarizabilidade, o calor de formação, as constantes de ionização ácida pKa, e assim por diante. É comum também o cálculo de descritores chamados químico-quânticos, extraídos diretamente dos cálculos quânticos tais como energia dos orbitais HOMO e LUMO, a diferença de energia entre eles, dentre outros parâmetros, mas é importante ressaltar que apenas estimativas teóricas desses existem, pois o correspondente experimental não é mensurável no laboratório.

Além desses descritores, representando a grande maioria dos descritores usados, porque são muito numerosos, há os descritores teóricos referentes à forma (morfológicos) e à constituição atômica das moléculas. Esses descritores assumem inúmeras classes, mas convém destacar aqui as mais conhecidas, dentre as quais boa parte foi utilizada para a construção do modelo QSAR da série de monoterpênicos em estudo. São elas:

1. Descritores constitucionais, são aqueles obtidos da composição da molécula (peso molecular, nº de átomos, nº de átomos de hidrogênio, nº de átomos "pesados, ou seja, que não o hidrogênio, nº de átomos de determinado

- elemento químico, soma dos volumes atômicos de van der Waals, nº de ligações simples, duplas etc., número de anéis, etc.);
2. Descritores topológicos: obtidos de grafos moleculares (teoria dos grafos) (geralmente do esqueleto da molécula, ou seja, sem hidrogênios), isto é, são descritores bidimensionais independentes da conformação (2D-descriptors);
 3. Descritores de Contagem de Caminhos e Caminhadas (*Walk and path counts*) (polaridade);
 4. Índices de conectividade (tipos de ligação, etc.);
 5. Descritores geométricos: diferentes tipos de descritores conformacionalmente dependentes baseados na geometria molecular. Valores confiáveis são obtidos se as conformações confiáveis foram previamente calculadas.
 6. Contagem de grupos funcionais: baseados na contagem de grupos funcionais químicos.
 7. Descritores de autocorrelação 2D: calculados a partir do grafo molecular somando os produtos dos pesos dos átomos dos átomos terminais de todos os caminhos do comprimento do caminho considerado.
 8. Descritores 3D Morse: calculados pela soma dos pesos dos átomos tomados por uma função de espalhamento angular diferente.

2.5 MODELOS QSAR BASEADOS EM REGRESSÃO LINEAR MÚLTIPLA

Em Ciência, de forma geral, é comum a necessidade de encontrar relações quantitativas entre um fenômeno observado e algumas variáveis independentes que se acreditam ter relevância na explicação do fenômeno. Em modelagem QSAR, o fenômeno observado é a atividade biológica e as variáveis independentes são propriedades físico-químicas ou parâmetros numéricos que sintetizem as diferentes informações químicas referentes a cada composto (descritores moleculares, seção anterior). A Regressão Linear Múltipla (MLR) é um modelo matemático que busca averiguar uma possível relação quantitativa linear entre a atividade (função resposta, Y) e as variáveis independentes do problema, os descritores moleculares (X_N).

O modelo linear reside em escrever a função resposta como uma combinação linear de variáveis independentes, também chamadas explicativas, X_1, X_2, \dots, X_k , capaz de reproduzir da melhor forma possível os valores experimentais de um grupo de N observações do fenômeno. Na equação 6, é mostrada a combinação linear de Y em relação às variáveis X_N .

Equação 6 - Modelo Ideal de Regressão

$$Y = A_0 + A_1X_1 + A_2X_2 + A_3X_3 + \dots + A_NX_N + E$$

Fonte: (BELEM,2016).

Em que:

“Y” é a função resposta, relacionada diretamente à atividade biológica;

X_N são as variáveis (descritores);

Os coeficientes “ A_1, A_2, \dots, A_N ” são considerados parâmetros do modelo;

“N” é o número de variáveis no modelo;

“E” é o erro explícito do modelo;

O coeficiente “ A_0 ” é a constante de ajuste.

O sucesso da aplicação do método de Regressão Linear Múltipla para um modelo QSAR pode ser avaliado por diversos indicadores estatísticos. Essa parte da avaliação é feita a partir do cálculo do coeficiente de correlação (R), e do desvio-padrão (s), e do fator F. A avaliação do ajuste do modelo pode ser feita pela análise da variância da regressão. O que se espera de um modelo em relação ao grau de ajuste é que ele apresente R o mais próximo possível de 1, que o valor de s seja o mais próximo possível de zero e que o número F seja o mais alto possível, baseado num referencial previamente estimado com determinado grau de confiança.

A qualidade do modelo final na capacidade de previsão depende da qualidade dos dados experimentais de atividade e dos descritores obtidos, além de aspectos estatísticos e, portanto, pode ser capaz de explicar o fenômeno observado, quando os descritores corresponderem a propriedades físico-químicas mensuráveis, mas também de proporcionar previsões de atividade para compostos ainda não testados no laboratório.

2.6 MODELOS QSAR BASEADOS EM OUTROS MÉTODOS DE *MACHINE LEARNING*

Neste estudo, usamos três métodos de aprendizagem de máquina na plataforma OCHEM, a MLR, o PLS e a NN. Os dois últimos métodos, Mínimos Quadrados Parciais (PLS – *Partial Least Squares*) e Redes neurais (NN - *Neural networks*), não serão discutidos aqui em detalhes, especialmente o PLS, devido a seu formalismo estar fora do escopo da monografia.

“Redes neurais é o nome dado a um conjunto de métodos matemáticos e algoritmos computacionais especialmente projetados para simular o processamento de informações e aquisição de conhecimento do cérebro humano” (CERQUEIRA, 2001). Redes neurais (NN) são um subgrupo de algoritmos de aprendizagem de máquina, que foi idealizado em 1943 e, em 1959 já tinha sua primeira aplicação em um filtro para eliminar ecos em linhas telefônicas. Modernamente, com a evolução da

Ciência e Tecnologia computacional nas últimas décadas, as NN fizeram enorme sucesso e há diversas aplicações, como jogar jogos complexos, analisar rostos e identificar a pessoa, analisar e copiar voz, entre outras aplicações que não seriam possíveis de implementar com programação comum, por causa da complexidade do problema (OLIVEIRA-FILHO, 2021).

Para este trabalho é suficiente dizer que o método PLS (que é uma modificação do método de Análise de Componentes Principais – PCA) é um método multivariado que, ao contrário do método MLR, pode ser aplicado a conjuntos menores de dados, sendo esta sua vantagem e um dos motivos pelo qual foi escolhido na presente monografia. O método das Redes Neurais (NN), de forma complementar ao PLS, supre a deficiência do método MLR quando se tem uma relação da atividade com os dados moleculares com elevado grau de não linearidade.

3 METODOLOGIA

3.1 OBTENÇÃO DADOS EXPERIMENTAIS DE CONTATO RESIDUAL

Todos os dados experimentais foram obtidos a partir de uma parceria entre o LAQTEO/UFRPE (Laboratório de Química Teórica - UFRPE) e o Laboratório de Produtos Naturais Bioativos – UFRPE, a partir do grupo de pesquisa do professor Dr. Cláudio Câmara. Em particular, os dados da atividade acaricida da série de monoterpenos utilizada nesta monografia foram produzidos no trabalho de Moraes (2012). A seguir é descrito um breve informativo sobre o procedimento experimental empregado por Moraes para obtenção das concentrações letais médias (CL₅₀) por contato Residual.

Todos os procedimentos experimentais foram administrados em fêmeas adultas do *Tetranychus urticae* (ácaro rajado), criadas em plantas de feijão-de-porco, *Canavalia ensiformes* no Laboratório de Biologia de Insetos do Departamento de Agronomia (UFRPE), sob condições de 26 ± 1°C, U.R. 65% ± 10% e fotofase de 12 horas. Todos os insetos usados foram originários de criação.

Os compostos puros foram comprados da Sigma-Aldrich® (Buenos Aires, Argentina), os compostos com atividades testadas para o ácaro rajado, utilizados neste trabalho foram: α -pineno, *p*-cimeno, β -pineno, mentol, α -terpineol, geraniol, R-limoneno, 1-8-cineole, R-linalol, acetato de terpenila, terpinoleno e 4-(S) terpineol. Todos com altíssima pureza (>90%). Os compostos foram escolhidos por apresentarem elevada similaridade química (>0,7), sendo todos monoterpenos ou derivados. A atividade inseticida utilizada neste trabalho foi obtida por meio de contato residual.

3.1 PROCEDIMENTO TEÓRICO

3.2.1 Determinação da estrutura molecular

As estruturas moleculares foram determinadas por métodos de otimização de geometria via métodos químico-quânticos para estados de mínimo de energia, primeiramente por métodos semiempíricos e refinados via métodos *ab initio* e/ou DFT (Teoria do Funcional da Densidade) utilizando o programa Gaussian 2009 (G09), em particular com o método/base B3LYP/6-31+G(d). Não foi realizada uma análise conformacional dos compostos calculados, no vácuo, ficando como perspectivas de continuidade do trabalho.

3.2.2 Determinação dos descritores moleculares

3.2.2.1 Série de treino e série de teste:

De início, a série de compostos em estudo foi subdividida em duas categorias: (i) série de treino e (ii) série de teste. A série de treino é o conjunto de compostos (e respectivos valores de atividade) que alimentarão a plataforma para, numa primeira etapa calcular-se os descritores moleculares e em seguida a obtenção de modelos QSAR. A qualidade do modelo obtido depende de vários fatores: o grau de similaridade química entre os compostos, a boa qualidade dos dados experimentais, uma quantidade relativamente grande de compostos na série de treino (o que não é o caso do presente trabalho), uma boa escolha do conjunto de (tipos) descritores, um bom método e programa para o cálculo dos descritores e, por fim, um eficiente método de análise multivariada (“aprendizado de máquina” como a plataforma OCHEM designa). Costuma-se dizer que os compostos desta série irão “treinar” o modelo, por isso a expressão série de treino. Os compostos remanescentes, que ficam de fora da série de treino, serão utilizados para testar o modelo, após obtido, e constituirão a chamada série de teste.

Dos 12 monoterpenos, como atividade acaricida disponível tivemos que escolher 10 para série de treino, pois é o número mínimo que a OCHEM exige, ficando com apenas 2 compostos para teste.

3.2.2.2 Escolha dos descritores:

Ao mesmo tempo que foram calculadas as estruturas moleculares para cada um dos compostos, utilizando os softwares pelos métodos semiempíricos e DFT pelo Gaussian 2009, foram determinadas automaticamente suas estruturas eletrônicas e, portanto, os descritores teóricos químico-quânticos também. No entanto, o seu emprego para a obtenção dos modelos QSAR via a OCHEM não se mostrou útil até o momento, sendo superados pelos descritores obtidos automaticamente pela plataforma por programas *built-in* (embutidos). Nesta monografia nos limitaremos a apenas ilustrar aspectos da estrutura eletrônica dos compostos na seção de resultados, e serão plotados os orbitais moleculares de fronteira e suas respectivas energias.

Inicialmente, antes de se ter conhecimento da plataforma OCHEM, tentou-se utilizar o programa DRAGON na sua versão livre na internet, o e-DRAGON, que fora utilizado anos antes pelo ex-orientando do Prof. Hércio, José Lucas Fialho. No entanto, descobriu-se que esta versão foi descontinuada. Foi então na busca por soluções alternativas e de software livre, encontrou-se a plataforma on-line OCHEM, como opção viável.

Os descritores moleculares foram obtidos seguindo as etapas:

1. Confecção e *upload* (opção "*batch data upload*", ver Figura 4) na OCHEM de um arquivo no formato sdf, que contém as geometrias da série de treino no formato .mol em coordenadas internas em sequência. Este arquivo pode conter, como dado extra, de antemão, os valores das atividades acaricidas (CL₅₀). Outra forma de fazer upload é por meio de um arquivo excel com as mesmas informações, com mais alguns detalhes exigidos pela plataforma, mas de preferência já com os valores de atividade.
2. Escolha do conjunto de descritores, ou seja, escolhendo as diferentes opções e combinações oferecidas pela OCHEM, a qual inclui, de forma gratuita, a versão 7.0 do DRAGON, além de muitos outros conjuntos publicados na literatura.

É possível, após calculados os descritores, fazer o download dos em arquivos de formatos diferentes, em especial, .sdf ou de excel (XLSX).

3.2.2.3 Obtenção do modelo QSAR:

Os modelos foram obtidos seguindo as etapas:

3. Após calculados os descritores, é feita a escolha do método de “aprendizado de máquina” (método de análise multivariada), para o quais variamos entre 3 opções, a RLM, o PLS e a NN.
4. Escolha da forma matemática que transforma a função resposta (Y), que é o valor da atividade acaricida, para que a função auxiliar (Y') seja usada para o cálculo de correlação estatística com o conjunto de descritores calculados para o conjunto de treino. Quando o modelo é testado para um composto qualquer, o valor obtido Y' deve ser convertido, de forma inversa, para Y .

Os passos foram numerados em sequência, porque todos eles, na realidade desde o cálculo dos descritores até a obtenção do modelo, podem ser feitos de uma só vez na OCHEM. As etapas todas podem ser visualizadas na seção a seguir (figuras 2, 3, 4, 5, 6, 7 e 8), com a ilustração das telas consecutivas da plataforma que aparecem para o usuário durante o processo.

3.2.2.4 Ilustração passo-a-passo da interface OCHEM para o usuário durante o processo de obtenção do modelo QSAR:

Exibimos, como exemplo, o cálculo na OCHEM de um conjunto particular de descritores, Dragon versão 7, seguindo os passos de 1 a 2 das seções anteriores.

Figura 2. Imagem da plataforma web OCHEM na página inicial.

Online chemical database
with modeling environment

Home ▾ Database ▾ Models ▾

Welcome to OCHEM! Your possible actions

Explore OCHEM data
Search chemical and biological data: experimentally measured, published and exposed to public access by our users. You can also [upload your data](#).

Create QSAR models
Build QSAR models for predictions of chemical properties. The models can be based on the experimental data published in our database.

Run predictions
Apply one of the available models to predict property you are interested in for your set of compounds.

Screen compounds with ToxAlerts
Screen your compound libraries against structural alerts for such endpoints as mutagenicity, skin sensitization, aqueous toxicity, etc.

Optimise your molecules
Optimise different properties for your molecules (e.g., reduce their toxicity or improve their ADME properties) using the state-of-the-art MolOptimiser utility based on matched molecules.

Check out the properties available on OCHEM
OCHEM contains 3499558 records for 916 properties (with at least 50 records) collected from 15494 sources

Melting Point **logPow**
logBB LogL(water) **LogD** LogPI(+)

Water solubility LogL(blood)
LogL(oil) ER Cbrain/Cplasma **IC50**

Papp(Caco-2) Papp(MDCK)

Oral absorption LIC 50

Papp ratio(Caco-2)

Plasma protein binding
Papp ratio(MDCK-mdr1) **pIC50**

%Human FA **Human IA** Human FA

fraction unbound (fu)
fraction ionized (fi) **pKa** **VDss**

LogIC50 **LogPI**

Fonte: <https://ochem.eu/home/show.do>

Figura 3. Escolhendo no menu “Models” a opção “Calcular descritores”.

Online chemical database
with modeling environment

Home ▾ Database ▾ Models ▾

Welcome to OCHEM! Your possible actions

Explore OCHEM data
Search chemical and biological data: experimentally measured, published and exposed to public access by our users. You can also [upload your data](#).

Create QSAR models
Build QSAR models for predictions of chemical properties. The models can be based on the experimental data published in our database.

Run predictions
Apply one of the available models to predict property you are interested in for your set of compounds.

Screen compounds with ToxAlerts
Screen your compound libraries against structural alerts for such endpoints as mutagenicity, skin sensitization, aqueous toxicity, etc.

Optimise your molecules
Optimise different properties for your molecules (e.g., reduce their toxicity or improve their ADME properties) using the state-of-the-art MolOptimiser utility based on matched molecules.

Check out the properties available on OCHEM
OCHEM contains 3499558 records for 916 properties (with at least 50 records) collected from 15494 sources

Melting Point **logPow**
logBB LogL(water) **LogD** LogPI(+)

Water solubility LogL(blood)
LogL(oil) ER Cbrain/Cplasma **IC50**

Papp(Caco-2) Papp(MDCK)

Oral absorption LIC 50

Papp ratio(Caco-2)

Plasma protein binding
Papp ratio(MDCK-mdr1) **pIC50**

%Human FA **Human IA** Human FA

fraction unbound (fu)
fraction ionized (fi) **pKa** **VDss**

LogIC50 **LogPI**


Models ▾

- Create a model
 - Apply a model
 - Create multiple models
- Open predictor
 - Upload a linear model
 - Upload a stub model
 - Add model template
- View pending tasks
- View published tasks
- SetCompare utility
- MolOptimiser
- Calculate descriptors**
- Descriptors storage

<https://ochem.eu/descriptorscalculator/show.do>

Fonte: <https://ochem.eu/home/show.do>

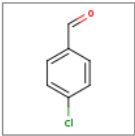
Figura 4. Fazendo o upload do arquivo com os dados da série de treino.


DescriptorsCalculator: Provide the molecules
 Provide the molecules, choose the descriptor types and download the descriptor values as an Excel, CSV or SDF file

Provide the compounds to calculate molecular descriptors for:

Upload compounds from a file
 SDF, MOL2, SMILES or an Excel sheet
 Nenhum arquivo escolhido

Draw Molecule
 click on depiction to the right to draw




Name/CASRN/SMILES:
 e.g., "CC=CCC" or "Aspirine"

Choose a previously prepared set: [...]

Select molecules by a tag: [...]

Fonte: <https://ochem.eu/home/show.do>

Figura 5. Pré-seleção de opções padrão da OCHEM (recomendadas).


DescriptorsCalculator: Structure preprocessing
 Select the structure preprocessing options

Preprocessing of molecules ⓘ

Choose your preferred standardizer: CHEMAXON ▾
 (Note: CDK standardization is currently experimental)

- Standardization
- Neutralize
- Remove salts
- Clean structure

Fonte: <https://ochem.eu/home/show.do>

Figura 6. Escolha dos descritores, no caso os da versão 7 do DRAGON.

The screenshot shows the 'Online chemical database with modeling environment' interface. The main heading is 'DescriptorsCalculator: Select the descriptors for comparison'. Below this, there is a section 'Please, select the descriptor types that you would like to calculate.' and a sub-section 'Select the molecular descriptors'. The interface is divided into several columns of checkboxes for different descriptor types:

- Recommended descriptor types (2D):** Includes OESate, ALogPS (2), Mol02 (777), CDDD, JPllogP, SIRMS, ISIDA Fragments, The in Hashed Atom Pair fingerprint (MAP4), OSFfragment (1138), QNPR, Multilevel Neighborhoods of Atoms (MNA), and Structural alerts (ToxAlerts and Functional Groups).
- Recommended descriptor types (3D):** Includes alvaDesc v. 2.0.4 (5866/3D) and Dragon v. 7 (5270/3D). A yellow highlight covers a large list of 3D descriptors, including Constitutional descriptors (47), Topological indices (75), Connectivity indices (37), 2D matrix-based descriptors (607), Burden eigenvalues (96), ETA indices (23), Geometrical descriptors (3D, 38), 3D autocorrelations (3D, 80), 3D-MoRSE descriptors (3D, 224), GETAWAY descriptors (3D, 273), Functional group counts (3D, 154), Atom-type E-state indices (172), 2D Atom Pairs (1596), Charge descriptors (3D, 15), Drug-like indices (28), Ring descriptors (32), Walk and path counts (46), Information indices (50), 2D autocorrelations (213), P_VSA-like descriptors (55), Edge adjacency indices (324), 3D matrix-based descriptors (3D, 99), RDF descriptors (3D, 210), WHIM descriptors (3D, 114), Randic molecular profiles (3D, 41), Atom-centred fragments (115), CATS 2D (150), 3D Atom Pairs (3D, 36), Molecular properties (20), and CATS 3D (3D, 300).
- Predictions by OCHEM's featured models:** Includes Ames Ievenberg, Toxicity against T. Pyriformis, ALogPS 3.0, CYP1A2 Estate+ALogPS, CYP2C9 Estate+ALogPS, CYP2C19 Estate+ALogPS, CYP2D6 Estate+ALogPS, CYP3A4 Estate+ALogPS, Pyrolysis point prediction (best Estate), Melting Point prediction (best Estate), Water solubility model based on logP and Melting Point, ALogPS 2.1 logP, and ALogPS 2.1 logS.
- Outputs of other OCHEM models:** Includes CDK 2.0 descriptors (256/3D), CDK 1.4.11 descriptors (256/3D), E-state, Dragon v. 5.4 (1644/3D), Dragon v. 5.5 (3224/3D), Dragon v. 6 (4885/3D), and MOPAC 7.1 descriptors (25/3D).
- Obsolete/Additional descriptor types:** Includes CDK 2.0 descriptors (256/3D), CDK 1.4.11 descriptors (256/3D), E-state, Dragon v. 5.4 (1644/3D), Dragon v. 5.5 (3224/3D), Dragon v. 6 (4885/3D), and MOPAC 7.1 descriptors (25/3D).

Fonte: <https://ochem.eu/home/show.do>

Na figura 6, estão selecionados os descritores do Dragon versão 7. Como descrito anteriormente, no decorrer do trabalho a escolha dos descritores pode ser divididas de acordo com sua natureza, descritores experimentais ou teóricos (CONSONNI; TODESCHINI 2010). Na figura 5 não temos esta opção, mas há uma forma de fazer *upload* de descritores calculados ou obtidos experimentalmente pelo usuário, que não mostraremos aqui.

Figura 7. Imagem após o procedimento dos cálculos.

The screenshot shows the 'DescriptorsCalculator: Calculation results' page. It features a summary table with the following data:

Total molecules:	10
Correctly calculated:	10
Errors:	0

Below the table is a green button labeled 'Proceed to the descriptors download page'.

Fonte: <https://ochem.eu/home/show.do>

Figura 8. Seleção dos dados e escolha do formato do documento a ser baixado.

The screenshot shows a web interface titled "Data export" with the subtitle "Export the selected data as an Excel, CSV or SDF file". Below the title, it says "Please, select the items that you want to export:". There are three links: "[select all]", "[select unrestricted only]", and "[select none]". A list of items follows, each with a checkbox:

- Structure (SMILES or SDF)
- CASRN
- RECORDID
- MOLECULEID
- External unique identifier
- Identifier in article (N)
- NAMES
- Introducers of the records
- Last modifiers of the records
- Publication IDs
- Error messages
- Conditions of experiments
- DESCRIPTORS
- Comments
- Inchi-key
- Merge information for the same molecule

At the bottom, there are four green buttons: "Get Excel file", "Get CSV file", "Get SDF file", and "Get R script".

Fonte: <https://ochem.eu/home/show.do>

Na figura 8, é feita a seleção dos dados e a escolha do formato do documento a ser baixado com os valores calculados dos descritores. Para a subsequente construção do modelo QSAR, após o cálculo dos descritores, é recomendado que o documento baixado esteja no formato de Excel (XLSX).

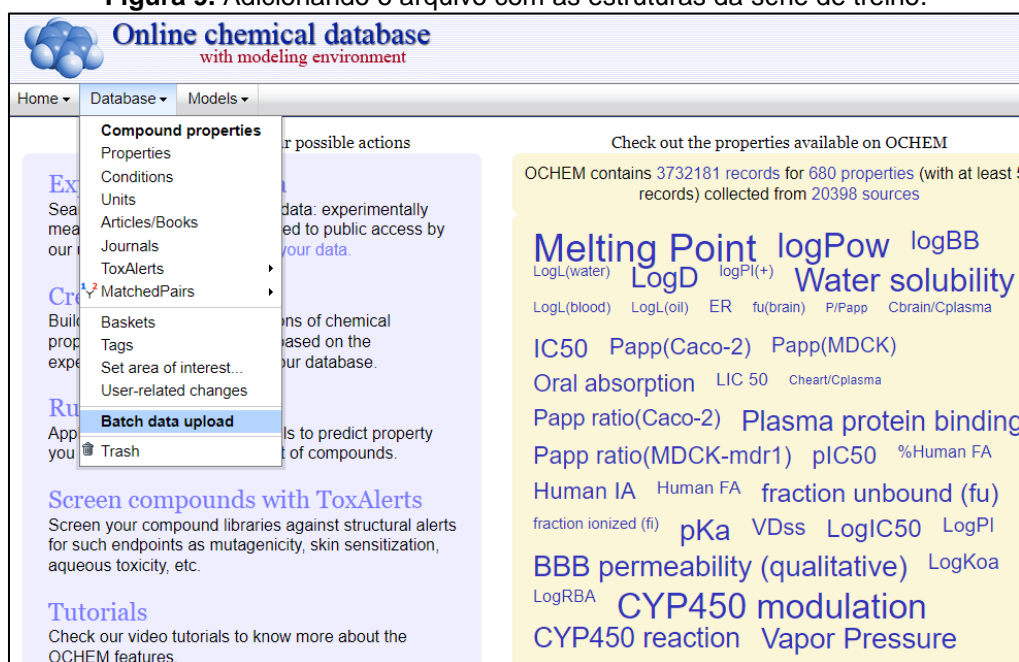
3.3 CONSTRUÇÃO E VALIDAÇÃO DOS MODELOS QSAR

Para a construção do modelo QSAR é obtido os cálculos dos descritores, que são suas variáveis, e possui os dados experimentais e teóricos dos compostos analisados. Por fim, para que possamos analisar esses modelos é necessário partir de vários critérios, sendo eles, a análise do coeficiente de correlação, o teste F, o valor-p, o nível geral de confiabilidade do modelo, e a correlação entre atividade encontrada e atividade calculada, podemos validá-lo pelo método clássico ou por validação cruzada, por exemplo (GAUDIO E ZANDONADE 2001).

No decorrer deste trabalho dividimos os compostos em duas série, a *série de teste*, compostas por dez compostos α -pineno, p-cimeno, β -pineno, α -terpineol, geraniol, r-limoneno, 1-8-cineole, r-linalol, acetato de terpenila e 4-(s) terpineol, e a *série de teste*, com 2 compostos mentol e terpinoleno. Totalizando 12 compostos analisados.

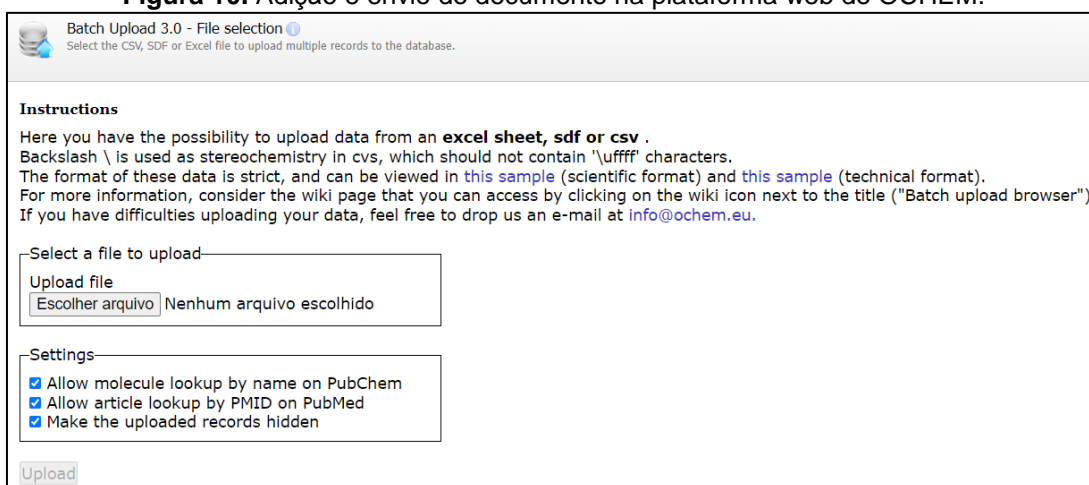
Antes da construção do modelo é necessário adicionar o documento contendo os dados dos nomes de cada composto, sua identidade e respectivamente sua atividade na plataforma on-line OCHEM. Esse documento deve estar no formato de Excel (XLSX), contendo todas essas informações. Nas figuras 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 e 20 é demonstrado como ocorre a adição deste documento e a construção do modelo para QSAR na plataforma.

Figura 9. Adicionando o arquivo com as estruturas da série de treino.



Fonte: <https://ochem.eu/home/show.do>

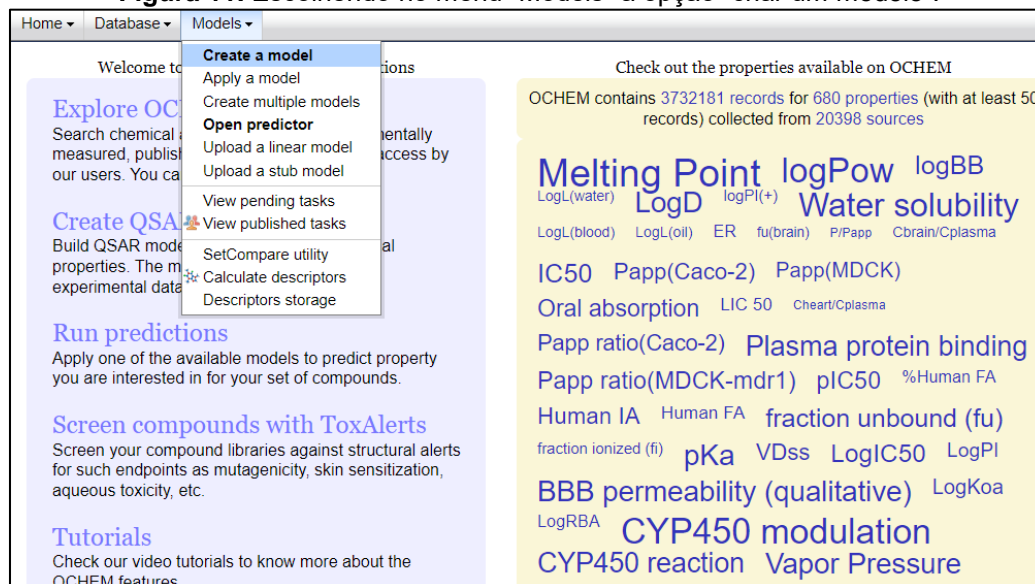
Figura 10. Adição e envio do documento na plataforma web do OCHEM.



Fonte: <https://ochem.eu/home/show.do>

Após a adição do documento na plataforma, é iniciado o procedimento para cálculo e construção do modelo.

Figura 11. Escolhendo no menu “Models” a opção “criar um modelo”.



Fonte: <https://ochem.eu/home/show.do>

Figura 12. Adição/validação dos descritores e escolha dos métodos para a construção do modelo.

Create a model ⓘ
Select the training and validation sets, the machine learning method and the validation protocol

Select the training and validation sets:

Training set (required): [...] [Add a validation set](#)

Choose the learning method: ⓘ

Suggested modeling methods:

- ASNN: ASsociative Neural Networks doi:10.1007/978-1-60327-101-1_10
- (New) Attentive FP doi: 10.1021/acs.jmedchem.9b00959
- ChemProp MPNN for property prediction (GPU) doi:10.1021/acs.jcim.9b00237
- CNF - Convolutional Neural Network Fingerprint (GPU) doi:10.1007/978-3-030-30493-5_79
- Transformer-CNF model
- Consensus model (based on models developed for the same set)
- DEEPCHEM: several methods from DeepChem (GPU) arXiv:1703.00564
- (New) DIMENET - Directional Message Passing Neural Network arXiv:2003.03123
- Deep Learning Consensus Architecture (DLCA) doi:10.1021/acs.jcim.9b00526
- DNN: Deep Neural Network (GPU) doi:10.1021/acs.jcim.8b00685
- EAGCNG - Edge Attention based Multi-relational Graph Convolutional Networks (GPU) arXiv:1802.04944
- FSMLR: Fast Stagewise Multiple Linear Regression doi:10.1134/S0012500807120026
- GNN - Graph Isomorphism Network (GPU) arXiv:1910.13124
- KNN: k - Nearest Neighbors

Fonte: <https://ochem.eu/home/show.do>

Figura 16. Seleção dos descritores para construção do modelo.

Model creator
Select model template and training set

Select the molecular descriptors ⓘ

<p>Recommended descriptor types (2D)</p> <ul style="list-style-type: none"> <input checked="" type="checkbox"/> OEState <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Bonds Indices <input type="checkbox"/> Counts only <input checked="" type="checkbox"/> ALogPS (2) <input type="checkbox"/> Mold2 (777) <input type="checkbox"/> CDDD <input type="checkbox"/> JPIogP <input type="checkbox"/> SIRMS <input type="checkbox"/> ISIDA fragments <input type="checkbox"/> The in Hashed Atom Pair fingerprint (MAP4) <input type="checkbox"/> GSFfragment (1138) <input type="checkbox"/> QNPR <input type="checkbox"/> Multilevel Neighborhoods of Atoms (MNA) <input type="checkbox"/> Structural alerts (ToxAlerts and Functional Groups) <p>Recommended descriptor types (3D)</p> <ul style="list-style-type: none"> <input type="checkbox"/> alvaDesc v.2.0.4 (5666/3D) 	<p>Predictions by OCHEM's featured models ⓘ</p> <ul style="list-style-type: none"> <input type="checkbox"/> Ames levenberg <input type="checkbox"/> Toxicity against T. Pyriformis <input type="checkbox"/> ALogPS 3.0 <input type="checkbox"/> CYP1A2 Estate+ALogPS <input type="checkbox"/> CYP2C9 Estate+ALogPS <input type="checkbox"/> CYP2C19 Estate+ALogPS <input type="checkbox"/> CYP2D6 Estate+ALogPS <input type="checkbox"/> CYP3A4 Estate+ALogPS <input type="checkbox"/> Pyrolysis point prediction (best Estate) <input type="checkbox"/> Melting Point prediction (best Estate) <input type="checkbox"/> Water solubility model based on logP and Melting Point <input type="checkbox"/> ALOGPS 2.1 logP <input type="checkbox"/> ALOGPS 2.1 logS <p><input type="checkbox"/> Outputs of other OCHEM models</p> <p>Obsolete/Additional descriptor types</p> <ul style="list-style-type: none"> <input type="checkbox"/> CDK 2.0 descriptors (256/3D)
---	--

Fonte: <https://ochem.eu/home/show.do>

Figura 17. Continuação da seleção dos descritores para construção do modelo.

<p>Recommended descriptor types (3D)</p> <ul style="list-style-type: none"> <input type="checkbox"/> alvaDesc v.2.0.4 (5666/3D) <input type="checkbox"/> Dragon v. 7 (5270/3D) <input type="checkbox"/> CDK 2.7.1 descriptors (256/3D) <input type="checkbox"/> Chemaxon descriptors (499/3D) <input type="checkbox"/> RDKit descriptors (3D) <input type="checkbox"/> MORDRED descriptors (1826/3D) <input type="checkbox"/> MOPAC2016 descriptors (35/3D) <input type="checkbox"/> KrakenX descriptors (MOPAC2016 derived)(124/3D) <input type="checkbox"/> PyDescriptor descriptors (16251/3D) <input type="checkbox"/> MERA descriptors (529/3D) <input type="checkbox"/> MERSY descriptors (42/3D) <input type="checkbox"/> 'Inductive' descriptors (54/3D) <input type="checkbox"/> Spectrophores (144/3D) 	<p>Obsolete/Additional descriptor types</p> <ul style="list-style-type: none"> <input type="checkbox"/> CDK 2.0 descriptors (256/3D) <input type="checkbox"/> CDK 1.4.11 descriptors (256/3D) <input type="checkbox"/> E-state <input type="checkbox"/> Dragon v. 5.4 (1644/3D) <input type="checkbox"/> Dragon v. 5.5 (3224/3D) <input type="checkbox"/> Dragon v. 6 (4885/3D) <input type="checkbox"/> MOPAC 7.1 descriptors (25/3D)
<p>Special descriptors (scaffolds, fingerprints):</p> <ul style="list-style-type: none"> <input type="checkbox"/> Chemaxon Scaffolds <input type="checkbox"/> Silicos-It Scaffolds <input type="checkbox"/> ECFP Fingerprints <input type="checkbox"/> MolPrint Fingerprints 	
<p>Under development: can change anytime and backward compatibility is not guaranteed. Use at your own risk!</p> <ul style="list-style-type: none"> <input type="checkbox"/> PaDEL2 descriptors (3D) <input type="checkbox"/> COSMO-RS SIGMA profiles (MOPAC2016 derived, (3D)) 	

Fonte: <https://ochem.eu/home/show.do>

Figura 18. Seleção para filtragem dos descritores.

Model creator
Select model template and training set

Select filters of descriptors

Eliminate descriptors with less than unique values

Delete descriptors that have absolute values larger than

Delete descriptors that have variance smaller than

Group descriptors, that have pair-wise correlations Pearson's correlation coefficient R larger than

Use Unsupervised Forward Selection to delete variables using the above value of multiple correlation coefficient R

After filtering, I want to select necessary descriptors myself (advanced)

Fonte: <https://ochem.eu/home/show.do>

Figura 19. Configuração para o modelo.

Model creator
Select model template and training set

Configure ANN method ⓘ

Training method:

Number of neurons in hidden layer:

Learning iterations (learning iterations):

Ensemble:

Enable ASNN:

Additional Parameters (separated by comma):

Experimental Parameters (conditions merging):

Fonte: <https://ochem.eu/home/show.do>

Figura 20. Início do cálculo para o modelo.

Model creator
Select model template and training set

Start calculation of the model

Now we are ready to start calculation.
Please provide the name for your model:

Save models

Task priority:

High priority (please, use for fast tasks only)

Normal priority

Low priority (for long tasks)

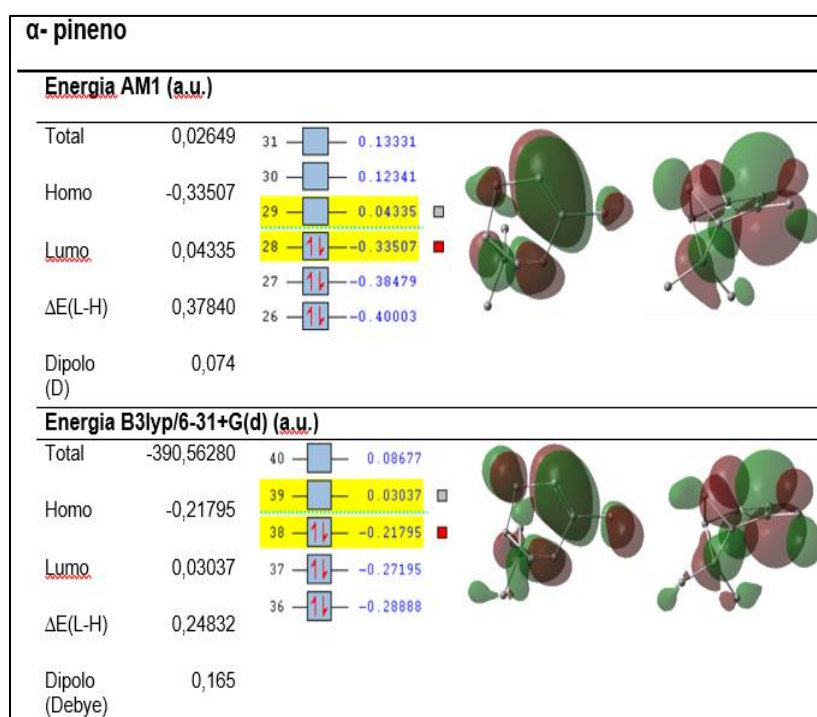
Fonte: <https://ochem.eu/home/show.do>

4 ESTRUTURA MOLECULAR E ELETRÔNICA DOS COMPOSTOS

Após a escolha dos compostos utilizados para série de treino e teste, foram realizadas as otimizações de geometria para obtenção das estruturas moleculares e eletrônicas de todos os 12 monoterpenos, por meio dos programas do GaussView 5.0 e Gaussian 09W. Primeiramente, as moléculas foram otimizadas pelos métodos AM1 e na sequência reotimizadas (refinados) pelo método DFT segundo o protocolo B3lyp/6-31+G (d). Como resultados obtivemos os seguintes dados: valores de energia total, energia dos orbitais HOMO (Orbital mais alto ocupado, do inglês “*Highest Occupied Molecular Orbital*”) e LUMO (Orbital mais baixo desocupado, do inglês “*Highest Occupied Molecular Orbital*”), momento de dipolo.

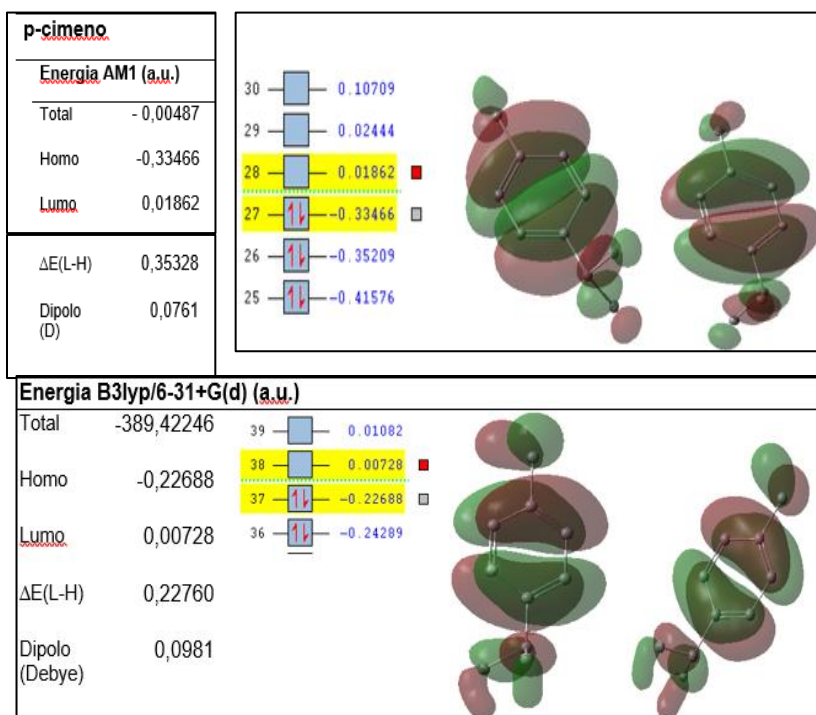
Nas páginas a seguir são apresentadas figuras (21 a 32) que ilustram os resultados para as energias dos orbitais de fronteira HOMO e LUMO, bem como uma representação gráfica da forma espacial dos mesmos, para todos os doze compostos estudados neste trabalho: α -pineno, p -cimeno, β -pineno, mentol, α -terpineol, geraniol, R-limoneno, 1-8-cineole, R-linalol, acetato de terpenila, terpinoleno e 4-(S) terpineol.

Figura 21. Energia e forma do orbitais HOMO e LUMO do composto α -pineno.



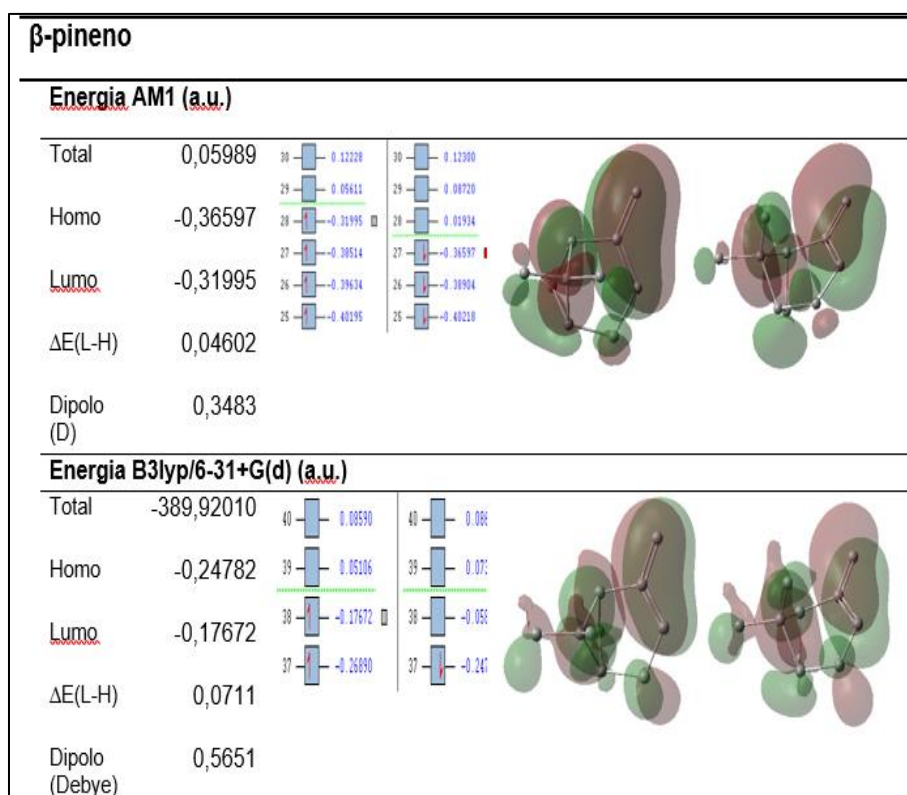
Fonte: Gaussian 09W.

Figura 22. Energia e forma do orbitais HOMO e LUMO do composto *p*-cimeno.



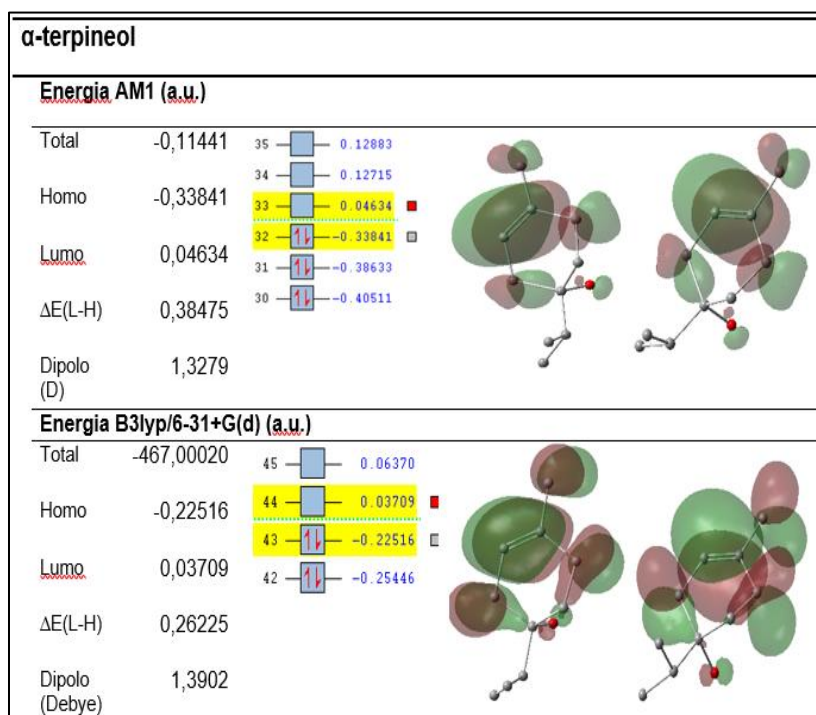
Fonte: Gaussian 09W.

Figura 23. Resultados da otimização dos compostos β -pineno.



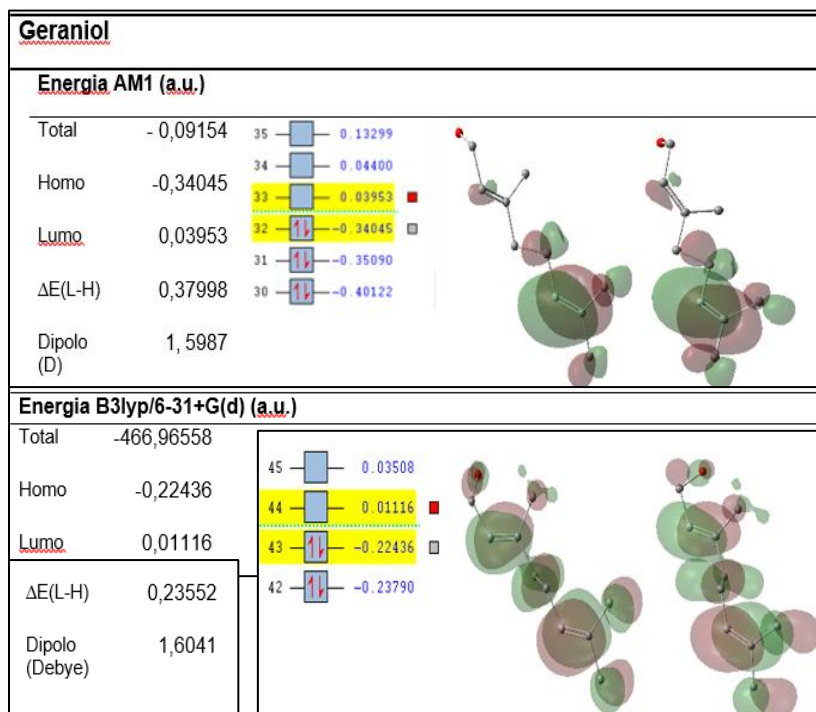
Fonte: Gaussian 09W.

Figura 24. Energia e forma do orbitais HOMO e LUMO do composto α -terpineol.



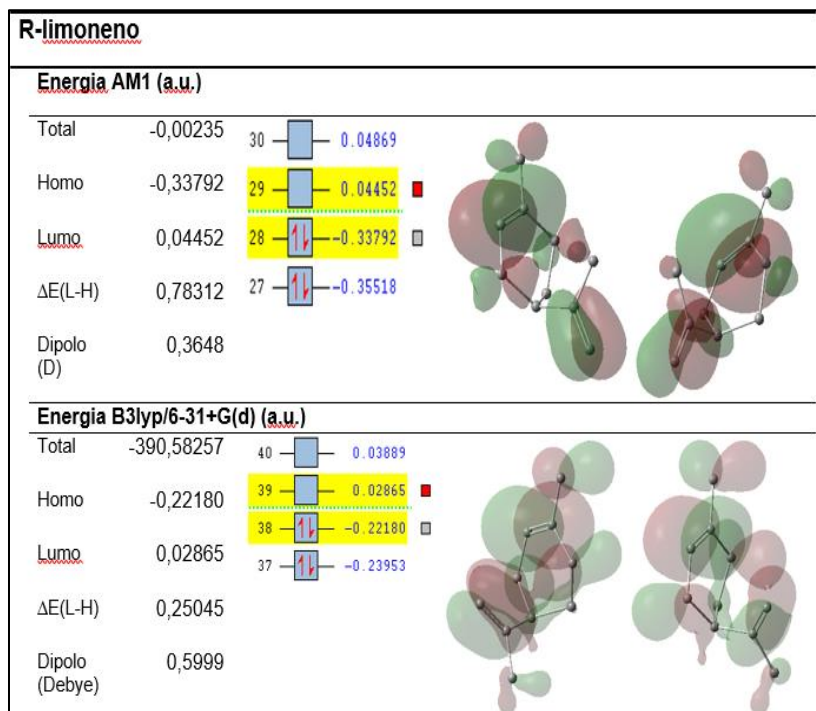
Fonte: Gaussian 09W.

Figura 25. Energia e forma do orbitais HOMO e LUMO do composto geraniol.



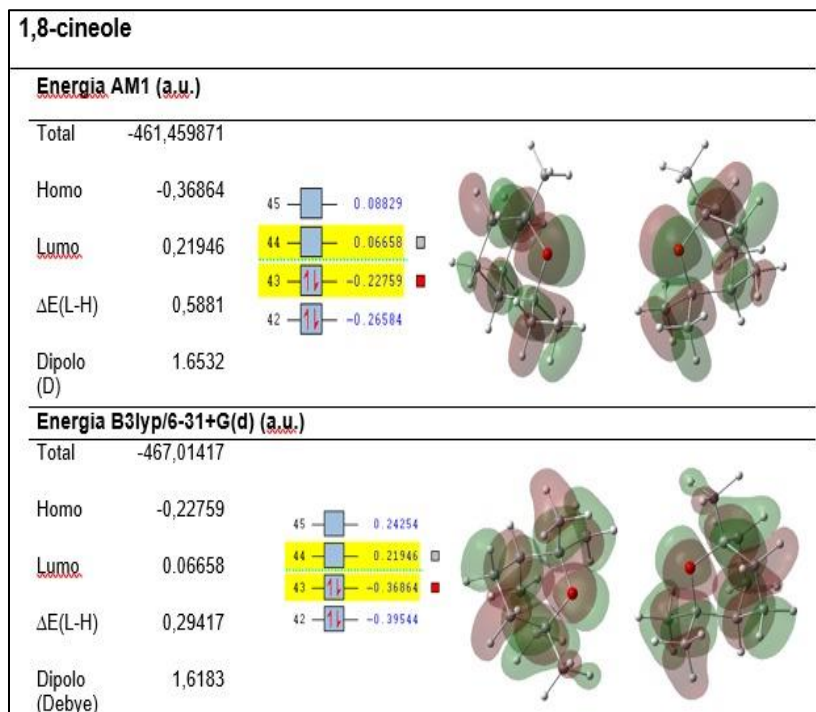
Fonte: Gaussian 09W.

Figura 26. Energia e forma do orbitais HOMO e LUMO do composto R-limoneno.



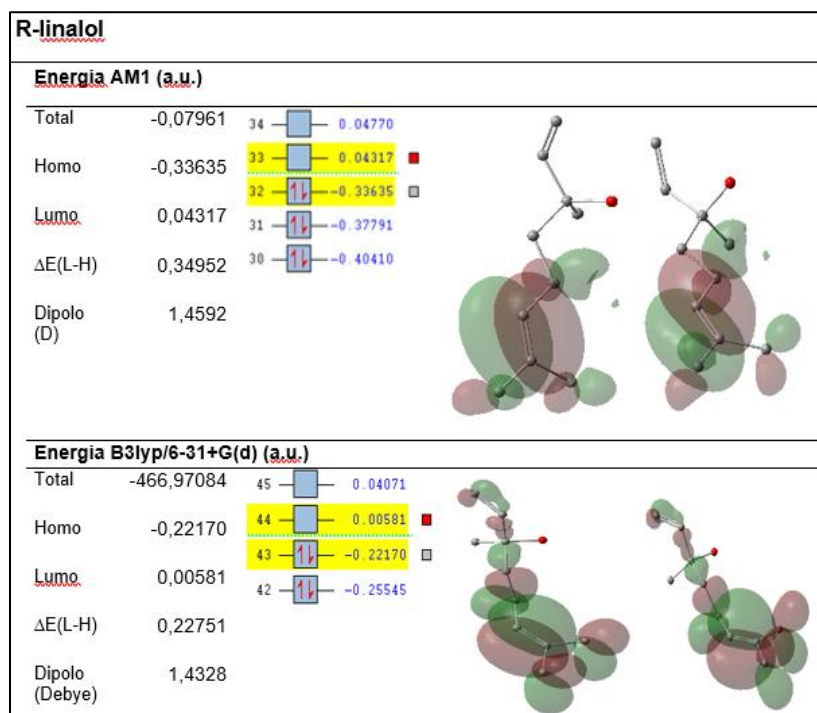
Fonte: Gaussian 09W.

Figura 27. Energia e forma do orbitais HOMO e LUMO do composto 1-8-cineole.



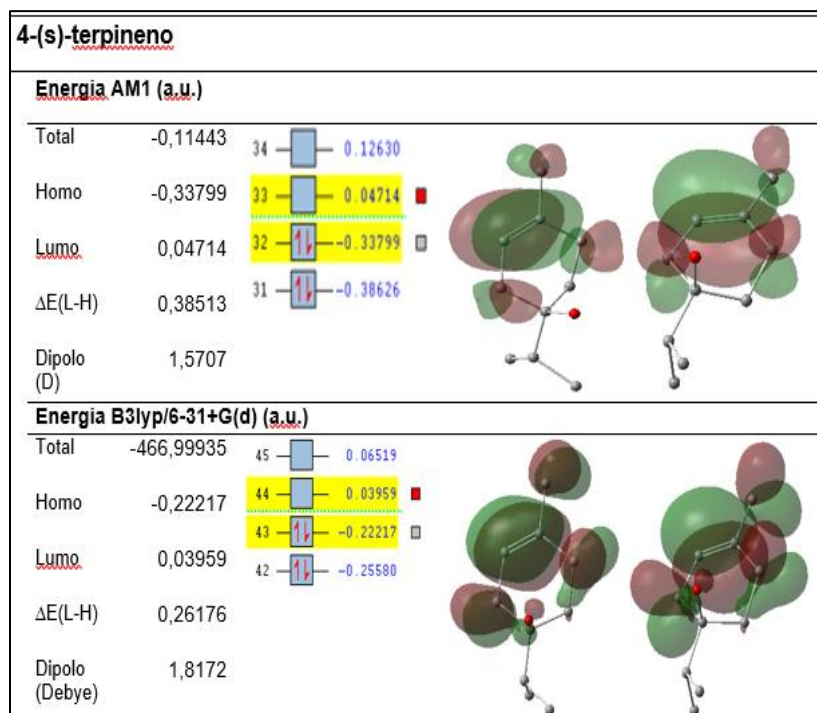
Fonte: Gaussian 09W.

Figura 28. Energia e forma do orbitais HOMO e LUMO do composto R-linalol.



Fonte: Gaussian 09W.

Figura 29. Energia e forma do orbitais HOMO e LUMO do composto 4-(S) terpineno.



Fonte: Gaussian 09W.

Figura 30. Energia e forma do orbitais HOMO e LUMO do composto mentol.

Mentol	
Energia AM1 (a.u.)	
Total	-0,12281844
Homo	-0,40179
Lumo	-0,29570
$\Delta E(L-H)$	0,10609
Dipolo (D)	1,4962
Energia B3lyp/6-31+G(d) (a.u.)	
Total	-467,564063
Homo	-0,29087
Lumo	-0,14414
$\Delta E(L-H)$	0,14673
Dipolo (Debye)	1,6496

Fonte: Gaussian 09W.

Figura 31. Energia e forma do orbitais HOMO e LUMO do composto acetato de terpenila.

Acetato de Terpenila	
Energia AM1 (a.u.)	
Total	-0,0221116
Homo	-0,34416
Lumo	0,31103
$\Delta E(L-H)$	0,03313
Dipolo (D)	1,5651
Energia B3lyp/6-31+G(d) (a.u.)	
Total	-620,12542
Homo	-0,23102
Lumo	-0,23112
$\Delta E(L-H)$	0,0001
Dipolo (Debye)	1,6970

Fonte: Gaussian 09W.

Figura 32. Energia e forma do orbitais HOMO e LUMO do composto terpinoleno.

Terpinoleno					
Energia AM1 (a.u.)					
Total	-0,014909				
Homo	0,04161				
Lumo	-0,32754	29	0,04161	□	
$\Delta E(L-H)$	-0,36915	28	-0,32754	■	
Dipolo (D)	0,1214				
Energia B3lyp/6-31+G(d) (a.u.)					
Total	-389,92010				
Homo	-0,24782				
Lumo	-0,17672	39	0,03451	□	
$\Delta E(L-H)$	0,0711	38	-0,21502	■	
Dipolo (Debye)	0,5651				

Fonte: Gaussian 09W.

Os cálculos de otimização de geometria foram convergidos satisfatoriamente quanto aos critérios numéricos, mas, em alguns casos, apesar da convergência numérica houve valores não apropriados para a constante de virial (valores menores que 2) referente à estrutura de mínimo, em particular para os compostos α -pineno e β -pineno pelo método AM1, caso em que mesmo a energia (calor de formação/AM1) resultou positiva, quando é esperado um valor negativo para compostos estáveis (CALVACANTI, 2016; ATKINS, 2012).

Uma observação desses resultados é que as energias dos orbitais HOMO são sempre negativas, o que é esperado em cálculos de moléculas estáveis, corroborando, portanto, as obtenções de mínimo de energia. No caso dos orbitais LUMO, em geral os valores foram positivos (indicando afinidades eletrônicas baixas ou negativas das moléculas calculadas), exceto para o composto β -pineno que apresenta o orbital LUMO negativo, indicando assim uma alta afinidade eletrônica (ATKINS, 2012).

É importante ressaltar que estas análises da afinidade eletrônica são em geral aproximações grosseiras para o caso de energias de orbitais, mas que funcionam razoavelmente bem para moléculas orgânicas (“bem-comportadas”) bastante

estáveis. Essas características dos orbitais HOMO e LUMO foram observadas para ambos os métodos AM1 e B3lyp/6-31+G(d) (ATKINS, 2012).

5 APLICAÇÃO DA OCHEM PARA OBTENÇÃO DE MODELOS QSAR PARA UMA SÉRIE DE MONOTERPENOS

5.1 Ilustração do arquivo de saída da OCHEM com os descritores físico-químicos

Conforme ilustra a Figura 8 da página 41, pode-se escolher a opção de download para um arquivo excel. A seguir mostramos, nas figuras 33, 34 e 35, partes de um arquivo com descritores calculados obtido a partir da plataforma OCHEM. Observamos que as estruturas da série de treino estão escritas num formato bastante sintético, muito comum em modelagem de fármacos, o SMILES. Observamos que na tabela constam apenas nove compostos na série de treino, mas na seção seguinte, da obtenção dos modelos QSAR, justificamos porque acabamos usando dez compostos para o treino.

Figura 33. Estruturas no formato SMILES de nove compostos usados na série de treino.

	A	B	C	D	E
1	SMILES	ARTICLEID	LC50TURC	NAME	
2	<chem>CC1=CC[C@@H]2C[C@H]1C2(C)C</chem>	A138179	34.13	a-pineno	
3	<chem>CC(C)C1=CC=C(C)C=C1</chem>	A138179	36.95	p-cimeno	
4	<chem>CC1(C)[C@H]2C[C@@H]1C(=C)CC2</chem>	A138179	38.93	b-pineno	
5	<chem>CC1=CC[C@H](CC1)C(C)(C)O</chem>	A138179	40.43	a-terpineol	
6	<chem>CC(C)=CCC\C(C)=C\CO</chem>	A138179	47.15	geraniol	
7	<chem>CC(=C)[C@@H]1CCC(C)=CC1</chem>	A138179	76.94	r-limoneno	
8	<chem>C[C@@]12CC[C@@H](CC1)C(C)(C)O2</chem>	A138179	91.80	1,8-cineole	
9	<chem>CC(C)=CCC[C@](C)(O)C=C</chem>	A138179	138.88	r-linalol	
10	<chem>CC(C)[C@]1(O)CCC(C)=CC1</chem>	A138179	366.56	4-s-terpineol	
11					
12					

Fonte: <https://ochem.eu/home/show.do>

Figura 34. Valores de alguns descritores para o Dragon versão 7, para as estruturas dos compostos calculadas pelo método AM1.

MW:(Dragon7)	AMW:(Dragon7)	Sv:(Dragon7)	Se:(Dragon7)	Sp:(Dragon7)	Si:(Dragon7)	Mv:(Dragon7)
136.0	5.24	14.2	25.1	16.1	29.3	0.547
134.0	5.59	13.7	23.2	15.3	26.9	0.57
136.0	5.24	14.2	25.1	16.1	29.3	0.547
154.0	5.32	15.5	28.3	17.3	32.9	0.533
154.0	5.32	15.5	28.3	17.3	32.9	0.533
136.0	5.24	14.2	25.1	16.1	29.3	0.547
154.0	5.32	15.5	28.3	17.3	32.9	0.533
154.0	5.32	15.5	28.3	17.3	32.9	0.533
154.0	5.32	15.5	28.3	17.3	32.9	0.533

Fonte: <https://ochem.eu/home/show.do>

Figura 35. Valores de alguns descritores para o Dragon versão 7, para as estruturas dos compostos calculadas pelo método B3LYP.

ALogPS_k	ALogPS_l	MW:(Dragon7)	AMW:(Dragon7)	Sv:(Dragon7)	Se:(Dragon7)	Sp:(Dragon7)	Si:(Dragon7)
3.66	-2.94	136.0	5.24	14.2	25.1	16.1	29.3
4.17	-3.91	134.0	5.59	13.7	23.2	15.3	26.9
3.94	-3.33	136.0	5.24	14.2	25.1	16.1	29.3
2.81	-1.79	154.0	5.32	15.5	28.3	17.3	32.9
2.89	-2.05	154.0	5.32	15.5	28.3	17.3	32.9
4.5	-2.47	136.0	5.24	14.2	25.1	16.1	29.3
3.36	-3.84	154.0	5.32	15.5	28.3	17.3	32.9
2.68	-2.51	154.0	5.32	15.5	28.3	17.3	32.9
2.81	-1.79	154.0	5.32	15.5	28.3	17.3	32.9

Fonte: <https://ochem.eu/home/show.do>

Com esses arquivos pode-se verificar os valores do mesmo conjunto de descritores obtidos para cada composto. É importante ressaltar que vários descritores, durante o processo de escolha (Figura 6, pág. 40), acabam dando informações químicas redundantes, mas há um algoritmo na OCHEM que faz uma filtragem reduzindo em muito o número de variáveis que efetivamente serão utilizadas no modelo.

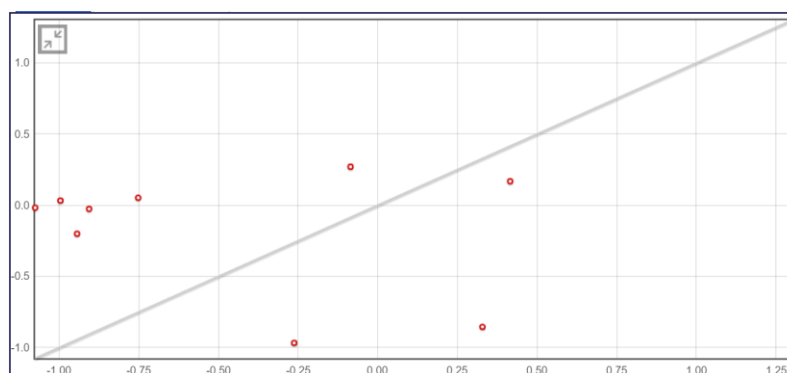
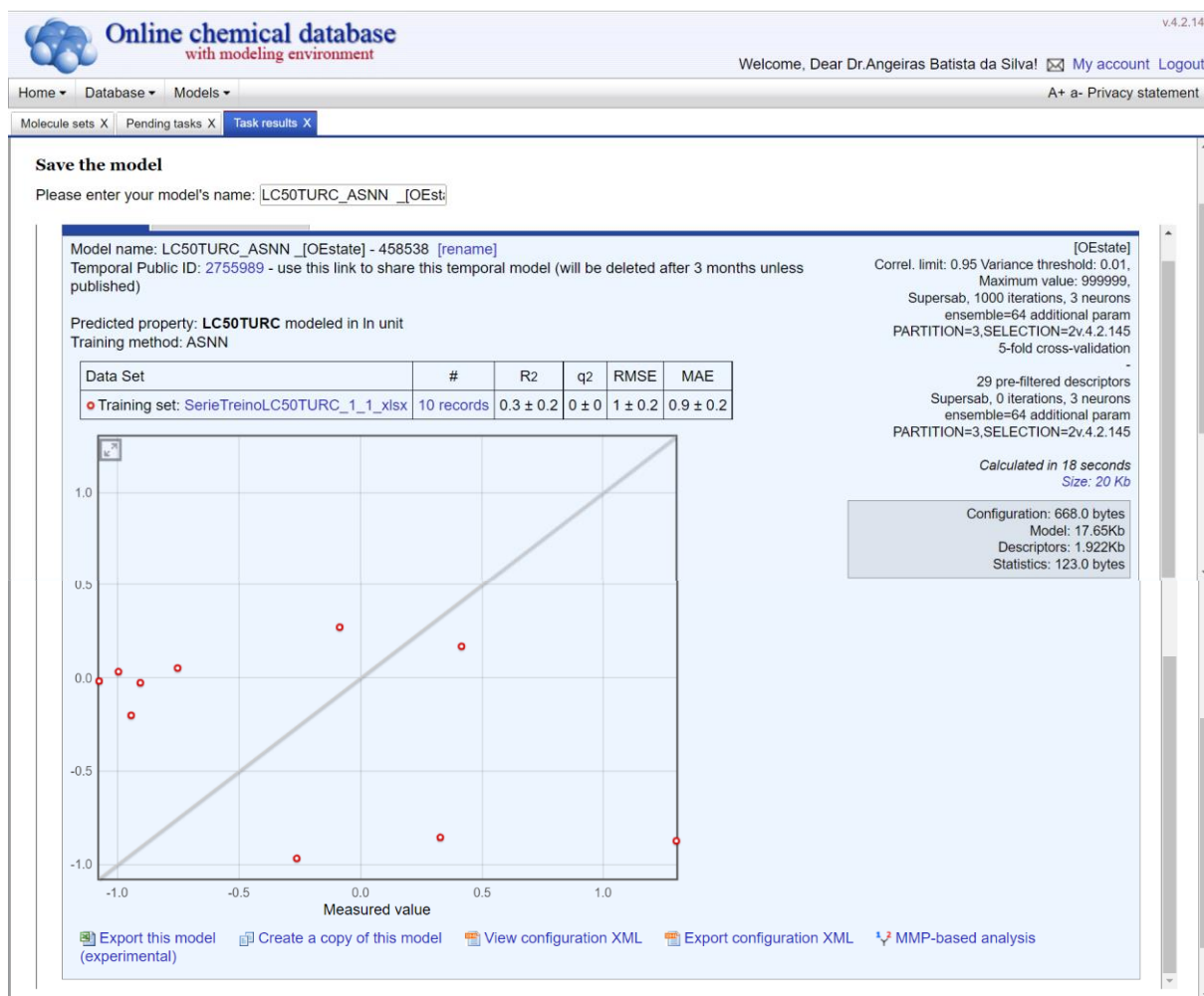
Na figura 6, a cada janela marcada, dentro da opção DRAGON v.7, temos subcategorias (subgrupos) de descritores e ao lado há um número que nos dá a quantidade de variáveis para aquela escolha específica. Na seção seguinte, nas figuras que ilustram o resultado final de um dos modelos calculados, aparece a quantidade total de variáveis usadas no modelo num quadro resumido, bem como a quantidade final de variáveis, após a filtragem devido a redundâncias.

5.2 Ilustração do arquivo de saída da OCHEM com alguns modelos QSAR

Importante lembrar que houve uma dificuldade inicial para obter o modelo com apenas nove compostos, para os quais se tinha calculado os descritores. Ou seja, de posse de doze compostos que apresentaram atividades inseticidas por contato residual contra o ácaro rajado (valores de CI50, listado na tabela 1), foram selecionados inicialmente apenas nove compostos para construção do modelo QSAR, sendo que os três compostos restantes serão utilizados como conjunto de teste. Mas ao tentamos construir o modelo QSAR, a plataforma não aceitou os dados, pois carece de um número mínimo de 10 compostos. Por esse motivo, adicionamos mais um composto ao conjunto de treino, o qual foi retirado da série de teste, que apresentou atividade mediana em comparação aos outros da mesma série, restando na mesma, para efeito de teste, dois compostos, um com baixa atividade e outro com alta atividade. Os compostos escolhidos para construção/tentativa do modelo foram: α -pineno, *p*-cimeno, β -pineno, α -terpineol, geraniol, R-limoneno, 1-8-cineole, R-linalol, acetato de terpenila, e 4-(S) terpineol.

Conforme se pode observar na seção de procedimento, as Figuras de 11 a 20, página 43 a 46, ilustram as opções para o cálculo do modelo. A seguir mostramos, na figura 36, um quadro resumo obtido da plataforma com informações das opções escolhidas pelo usuário para o cálculo do modelo, cujo gráfico resultante é também mostrado. Escolheu-se ilustrar um dos modelos obtidos que teve o maior valor de R^2 (quadrado do coeficiente de regressão linear), próximo de 0,3. Observamos que na figura constam os dez compostos da série de treino, o mínimo exigido pela plataforma.

Figura 36. Quadro resumo obtido da plataforma OCHEM com informações das opções escolhidas pelo usuário para o cálculo do modelo, com um gráfico ilustrando a qualidade do modelo, e com diversos indicadores estatísticos sobre a qualidade do modelo.



Fonte: <https://ochem.eu/home/show.do>

6 CONCLUSÃO E PERSPECTIVAS

A principal conclusão deste trabalho foi o sucesso na utilização dos recursos da plataforma livre OCHEM - *Online Chemical Modeling Environment* e, portanto, de experimentar modelagens computacionais em nuvem, para obtenção de modelos de QSAR. Conseguimos, ao longo do texto, ilustrar, com opções básicas da ferramenta, normalmente aquelas indicadas pelos autores, como alimentar o sistema com as estruturas dos compostos, como calcular diferentes descritores e como obter modelos QSAR. Foi possível fazer uma aplicação da OCHEM para obtenção de modelos QSAR para a atividade acaricida sobre o *Tetranychus urticae* (ácaro rajado) de uma série de monoterpenos: α -pineno, *p*-cimeno, β -pineno, mentol, α -terpineol, geraniol, R-limoneno, 1-8-cineole, R-linalol, acetado de terpenila, terpinoleno, 4-(S)-terpineol. Para este fim usamos 3 diferentes métodos de aprendizagem de máquina, quais sejam Regressão Linear Múltipla (MLR), Redes neurais (NN) e Mínimos Quadrados Parciais (PLS), com distintas formulações da função resposta (atividade acaricida), no entanto, os modelos obtidos para a série em estudo não tiveram qualidade estatística suficiente devido provavelmente, primeiro à quantidade muito pequena de compostos na série de treino do modelo com dados experimentais disponíveis e, também, ao provável grau de não linearidade presente na relação entre os descritores e a atividade. Para a realização destas ações fizemos: uma revisão bibliográfica e conceitual do método Relação Quantitativa Estrutura-Atividade, QSAR, do inglês *Quantitative Structure-Activity Relationship*; uma revisão bibliográfica do uso de monoterpenos no controle de pragas na agricultura e concluímos a modelagem molecular por métodos de Química quântica da série de monoterpenos em estudo, gerando as estruturas moleculares que alimentaram a plataforma OCHEM, para cálculo dos descritores moleculares e obtenção de modelos QSAR.

As perspectivas destacadas referem-se ao uso da plataforma para uma série de compostos bem maior em número com dados experimentais disponíveis e, assim, poder obter um modelo QSAR mais consistente e que possa auxiliar nas pesquisas do Laboratório de Produtos Naturais Bioativos da UFRPE por novos produtos ativos.

REFERÊNCIAS

- ABBOTT, T. P., ET. **“On-Line Chemical Database for New Crop Seeds”**. *Journal of the American Oil Chemists’ Society*, vol. 74, no 6, junho de 1997, p. 723–26. DOI.org (Crossref), <https://doi.org/10.1007/s11746-997-0208-8>.
- ANTUNES, A. Método de Hartree-Fock: Dois Exemplos Analiticamente Solúveis. *Revista Brasileira de Ensino de Física*, [s.l], v. 21, n. 2, p.221-232, jun. 1999. Disponível em: <www.sbfisica.org.br/rbef/pdf/v21_221.pdf>. Acesso em: 01 jul. 2021.
- ARROIO, Agnaldo; HONORIO, Káthia M.; SILVA, Albérico B. F. da. **Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade**. *Quím. Nova*, São Paulo, v. 33, n. 3, p. 694-699, 2010.
- ATKINS, P.; DE PAULA, J. **Atkins, física-química**. Rio de Janeiro: LTC, 2008. v. 2. ISBN 978- 85-216-1601-6.
- BARROS NETO, B.; De SCARMINIO, I. S.; BRUNS, R. E. **Planejamento e Otimização de Experimentos**, Ed. Unicamp, 1ª edição, 1995.
- BATISTA, Hélcio J. **Comunicação privada** (anotações de estudo dirigido durante o PIC UFRPE 2020/2021).
- BELEM, José L. F. **Estudo da Relação Quantitativa Estrutura-Atividade (QSAR) de Monoterpenos com Atividade Inseticida frente ao Ácaro Rajado (T. Urticae)**. 47 fls. Monografia de conclusão de Curso de Licenciatura em Química. 2016. UFRPE.
- BERLINCK, R. G. S.; BORGES, W. de S.; SCOTTI, M. T.; VIEIRA, P. C. **A Química de Produtos Naturais do Brasil do Século XXI**. *Química Nova*, Vol. 40, n. 6, 706-710, 2017.
- BERRIMAN, G. B.; DEELMAN, E.; JUVE, G.; RYNGE, M.; VÖCKLER, J.-S. **The application of cloud computing to scientific workflows: A study of cost and performance**. *Phil. Trans. R. Soc. A*. v. 371, 20120066, 2013.
- BEVERIDGE, M. (1985) **The development of young children's understanding of the process of evaporation** *British Journal of Educational Psychology* 55: 84 – 90
- BRAIBANTE, M. E. F.; ZAPPE, J. A. **A Química dos Agrotóxicos**. *Química Nova Na Escola*. Vol. 34, n.1, p. 10-15, 2012.
- da CAMARA, C. A. G.; LIMA, G. S.; MORAES, M. M.; Da SILVA, M. M. C.; MELO, J. P. R.; SANTOS, M. L.; FAGG, C. W. **Chemical composition and acaricidal activity of essential oils and selected terpenes from two species of *Psidium* in the Cerrado biome of Brazil against *Tetranychus urticae***. *Boletim Latinoamericano y del Caribe de Plantas Medicinales y Aromaticas* , v. 19, p. 15-28, 2020.
- CAVALCANTE, C. R. M. **Investigação de protocolos de cálculo por métodos de solvatação implícita na previsão do efeito gauche em 1,2-di-haloetanos**. 2016. 97 f. Dissertação (Mestrado) – Programa de pós-graduação em física, Universidade Federal Rural de Pernambuco, Recife, 2016.

CONSONNI, Viviana; TODESCHINI, Roberto. **MOLECULAR DESCRIPTORS**. In: PUZYN, Tomasz; LESZCZYNSKI, Jerzy; CRONIN, Mark T.d. (Ed.). Recent Advances in QSAR Studies: Methods and Applications. v 8. [s.l]: Springer, 2010. Cap. 3. p. 29-102.

CORREA, J.C.R.; SALGADO, H.R.N. **Atividade inseticida das plantas e aplicações: revisão**. Rev. bras. Plantas med., Botucatu, v. 13, n. 4, p. 500-506, 2011.

CRAMER, Bruno. Estudos de QSAR-2D **aplicados a diterpenóides clerodanos e dibenzoilidrazinas**. 2011. 316 f. Tese (Doutorado) - Curso de Programa de Pós-graduação em Química, Instituto de Química, Universidade Estadual de Campinas, Campinas, 2011. Cap. 5. Disponível em: <<http://www.bibliotecadigital.unicamp.br/document/?code=000818182>>. Acesso em: 26 dez. 2021.

de CERQUEIRA, E. O.; de ANDRADE, J. C.; POPPI, R. J.; MELLO, C. **Redes neurais e suas aplicações em calibração multivariada**. Quim. Nova, v. 24, n. 6, 864-873, 2001.

de OLIVEIRA-FILHO, A. G. S.; VICENTINI, E. D. **Introdução às redes neurais para regressões não-lineares: ajuste de superfícies de energia potencial**. Quim. Nova, v. 44, n. 2, 229-234, 2021.

DRAPER, N. R.; SMITH, H. **Applied Regression Analysis**. 3 ed. Nova Iorque: Wiley, 1998.

FERREIRA, Márcia Miguel Castro; MONTANARI, Carlos Alberto; GAUDIO, Anderson Coser. **Seleção de variáveis em QSAR**. Quím. Nova, São Paulo, v. 25, n. 3, p. 439-448, 2002.

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., JR.; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M.; HEYD, J. J.; BROTHERS, E.; KUDIN, K. N.; STAROVEROV, V. N.; KOBAYASHI, R.; NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A.; BURANT, J. C.; IYENGAR, S. S.; TOMASI, J.; COSSI, M.; REGA, N.; MILLAM, J. M.; KLENE, M.; KNOX, J. E.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; ZAKRZEWSKI, V. G.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; DAPPRICH, S.; DANIELS, A. D.; FARKAS, O.; FORESMAN, J. B.; ORTIZ, J. V.; CIOSLOWSKI, J.; FOX, D. J. **Gaussian 09 Revisão A.02**. Gaussian Inc. Wallingford CT 2009.

GAUDIO, ANDERSON COSER, E ELIANA ZANDONADE. **“Proposição, Validação e Análise Dos Modelos Que Correlacionam Estrutura Química e Atividade Biológica”**. Química Nova, vol. 24, no 5, outubro de 2001, p. 658–71. DOI.org (Crossref), <https://doi.org/10.1590/S0100-40422001000500013>.

HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA

IBM Research: **Science & Technology Outlook 2021. Accelerating discovery to solve our biggest challenges**. IBM Research. Janeiro de 2021. Disponível em <research.ibm.com>.

LÓPEZ, M.d.; PASCUAL-VILLALOBOS, M.j.. **Mode of inhibition of acetylcholinesterase by monoterpenoids and implications for pest control**. Industrial Crops And Products, [s.l.], v. 31, n. 2, p.284-288, mar. 2010. Elsevier BV. Disponível em: <http://www.sciencedirect.com/science/article/pii/S092666900900226X>. Acesso em: 20 abr. 2022.

MACHADO, C.; NICOLINIB, K. P.; NICOLINIB, J. **Estudos de QSAR para SARS-CoV-2**, Quim. Nova, v. 43, n. 9, 1258-1263, 2020.

MORAES, Marcilio Martins de. **Relação entre a toxicidade de mono e sequiterpenos identificados no óleo essencial de espécies do gênero Ocotea (Lauraceae) e Eugenia (Myrtaceae) sobre o ácaro rajado (tetranychus urticae)**. 2012. 111 f. Dissertação (Mestrado) - Programa de pós-graduação em química, Universidade Federal Rural de Pernambuco, Recife, 2012. Cap. 3.

MORGON, Nelson H.; CUSTÓDIO, Rogério. **Teoria do funcional da densidade**. Química Nova, [s.l.], v. 18, n. 18, p.44-55, jan. 1995. Disponível em: <http://quimicanova.sbq.org.br/imagebank/pdf/Vol18No1_44_v18_n1_10.pdf>. Acesso em: 20 abr. 2022.

NIST: **The NIST Definition of Cloud Computing - Recommendations of the National Institute of Standards and Technology**. NIST Special Publication 800-145. p. 2, 2011. Disponível em: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf> (Acessado em 06 de junho de 2022 às 08h00min).

OCHEM: OCHEM Docs. (*website*): <https://docs.ochem.eu/x/2QFr.html.>

POPLE, J.A AND BEVERIDGE, DL., **Approximate Molecular Orbital Theory**, McGraw, New York, 1969.

RODRIGUES, C. R. **Processos Modernos no Desenvolvimento de Fármacos: Modelagem Molecular**. Cadernos Temáticos de Química Nova na Escola, 2001.

SUSHKO, I.; NOVOTARSKYI, S.; KÖRNER, R.; PANDEY, A. K.; RUPP, M.; TEETZ, W.; BRANDMAIER, S.; ABDELAZIZ, A.; PROKOPENKO, V. V.; TANCHUK, V. Y.; TODESCHINI, R.; VARNEK, A.; MARCOU, G.; ERTL, P.; POTEMKIN, V.; GRISHINA, M.; GASTEIGER, J.; SCHWAB, C.; BASKIN, I. I.; PALYULIN, V. A.; RADCHENKO, E. V.; WELSH, W. J.; KHOLODOVYCH, V.; CHEKMAREV, D.; CHERKASOV, A.; AIRES-DE-SOUSA, J.; ZHANG, Q. Y.; BENDER, A.; NIGSCH, F.; PATINY, L.; WILLIAMS, A.; TKACHENKO, V.; TETKO, I. V. **Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information**. J Comput Aided Mol Des.; Vol. 25, n. 6, p. 533-554, 2011.

SZABO, A.; OSTILUM, N. S. **Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory**. 1 ed. Nova Iorque: Dover Publications, 1989.

TAVARES, Leoberto Costa. **QSAR: a abordagem de Hansch**. Quím. Nova, São Paulo, v. 27, n. 4, p. 631-639, 2004.

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**, J. Comput. Aid. Mol. Des., 2005, 19, 453-63, VCCLAB, Virtual Computational Chemistry Laboratory, <http://www.vcclab.org>, 2005. VCCLAB. **Virtual Computational Chemistry Laboratory. DRAGON Descriptors**. Website: http://www.vcclab.org/lab/indexhelp/dragon_descr.html.

VIEGAS, Junior, C. **Terpenos com atividade inseticida: uma alternativa para o controle químico de insetos**. Quím. Nova, São Paulo, v. 26, n. 3, p. 390-400, 2003.

WANG, M. W. H.; GOODMAN, J. M.; ALLEN, T. E. H. **Machine Learning in Predictive Toxicology: Recent Applications and Future Directions for Classification Models**. Chem. Res. Toxicol., v. 34, n. 2, p. 217–239, 2021.

ZHANG, P., TRUHLAR, D. G., GAO, J., Phys Chem Chem Phys. 14(21), 7821 (2012).

APÊNDICE

- Para uma lista completa dos descritores disponíveis na plataforma OCHEM e na página do DRAGON, com as respectivas referências bibliográficas dos grupos que os desenvolveram e explicação resumida do significado de muitos deles, e exemplos de como obtê-los, é importante consultar a documentação existente nas próprias páginas dos mesmos (OCHEM (*website*): <https://docs.ochem.eu/x/2QFr.html>); (VCCLAB. Virtual Computational Chemistry Laboratory. DRAGON Descriptors. Website: http://www.vcclab.org/lab/indexhelp/dragon_descr.html)).
- Para consultar uma lista organizada num só lugar é importante consultar o livro, “Molecular Descriptors for Chemoinformatics”, Volumes I e II, que traz uma lista em ordem alfabética de descritores, referências científicas e explicação resumida do significado de muitos deles, e exemplos de como obtê-los.