

Aplicação da Triagem Virtual no planejamento de teste colorimétrico para identificação preliminar de canabinoides sintéticos

R. da M. Macêdo ^{a,c*}, R. de F. Godoy ^a, K. S. da Costa ^b, W. S. de Alencar ^a

^a Programa de Pós-graduação em Ciências Forenses, Universidade Federal do Sul e Sudeste do Pará, Marabá, PA, Brasil

^b Universidade Federal do Oeste do Pará, Santarém, PA, Brasil

^c Laboratório Forense da Polícia Científica do Pará – Superintendência Regional II, Marabá, Pará, Brasil

*Endereço de e-mail para correspondência: rennedyfarmabio@gmail.com. Tel.: +55-99-981199556.

Recebido em 17/06/2024; Revisado em 06/08/2024; Aceito em 08/08/2024

Resumo

Os canabinóides sintéticos constituem um grupo de Novas Substâncias Psicoativas (NSP) que mimetizam os efeitos da *Cannabis sativa*, pois atuam nos mesmos receptores canabinoides. Atualmente representam o terceiro maior grupo de NSP. A constatação preliminar dessas drogas têm sido um verdadeiro desafio para os Laboratórios Forenses. Nesse contexto, o presente trabalho utiliza a técnica de Triagem Virtual baseada em ligantes para direcionar a seleção de moléculas com potencial atividade colorimétrica, aprimorando o processo de busca de um teste de cor eficiente e que contemplem a identificação da maioria de canabinoides sintéticos, além de tratar-se de uma alternativa que requer menores esforços financeiros, infraestrutura e tempo envolvido na descoberta. Para isso, foram selecionados o 4-dimetilaminobenzaldeído e o 2,4-dinitrofenilhidrazina, como compostos químicos de referência para filtragem molecular em banco de dados químicos, obtendo, respectivamente, 13 e 20 estruturas distintas entre si. Essas substâncias foram submetidas a metodologias de comparação por similaridade molecular: a *Principal Component Analysis* (PCA) e a *Cluster Analysis*, que utilizam características moleculares específicas dos compostos de referência, comparando-os com os correspondentes descritores das moléculas da base de dados, servindo-se de uma medida de similaridade química. Assim, foi possível apontar compostos com maior grau de similaridade aos compostos de referência, indicando candidatos promissores à identificação preliminar de canabinoides sintéticos.

Palavras-Chave: Canabinoides Sintéticos; Teste Colorimétrico; Triagem Virtual.

Abstract

Synthetic cannabinoids constitute a group of New Psychoactive Substances (NSP) that mimic the effects of *Cannabis sativa*, as they act on the same cannabinoid receptors. They currently represent the third largest group of NSPs. The preliminary detection of these drugs has been a real challenge for Forensic Laboratories. In this context, the present work uses the Virtual Screening technique based on ligands to direct the selection of molecules with potential colorimetric activity, improving the process of searching for an efficient color test that includes the identification of the majority of synthetic cannabinoids, in addition to This is an alternative that requires less financial effort, infrastructure and time involved in discovery. For this, 4-dimethylaminobenzaldehyde and 2,4-dinitrophenylhydrazine were selected as reference chemical compounds for molecular filtering in chemical databases, obtaining, respectively, 13 and 20 distinct structures. These substances were subjected to molecular similarity comparison methodologies: Principal Component Analysis (PCA) and Cluster Analysis, which use specific molecular characteristics of the reference compounds, comparing them with the corresponding descriptors of the base molecules. data, using chemical similarity measures. Thus, it was possible to identify compounds with a greater degree of similarity to the reference compounds, indicating promising candidates for the preliminary identification of synthetic cannabinoids.

Keywords: Synthetic Cannabinoides; Colorimetric test; Virtual Screening.

1. INTRODUÇÃO

Em uma situação de possível prisão em flagrante por posse ou tráfico de drogas, a Autoridade Policial deverá constatar a natureza e a quantidade da droga apreendida, conforme o Art. 50 e seu parágrafo 1º, da Lei nº 11.343 de 2006 (“Lei de drogas”). Portanto, a prisão em flagrante somente se sustenta com laudo de constatação da natureza e quantidade de droga apreendida, a ser elaborado em até 24 horas do flagrante. Nesse cenário, o teste preliminar se mostra válido até que efetivamente se possa proceder um exame definitivo, cientificamente inequívoco, que ateste a real natureza do material apreendido [1]. Percebe-se então, que apesar dos modernos laboratórios forenses empregarem a *Gas chromatography–mass spectrometry* (GC–MS) como a principal técnica analítica para identificação e confirmação de substâncias desconhecidas, ainda há um papel relevante para as técnicas de indicação presuntiva e constatação preliminar, como testes colorimétricos, sendo, no tocante às leis de caráter criminal, até o momento, o único tipo de laudo preliminar previsto para efeito de prisão em flagrante [2-3].

Os testes de cor fornecem indicação presuntiva da presença ou ausência de drogas específicas, ou mais precisamente determinadas classes estruturais de drogas ou grupos funcionais. Desta forma, facilitam a triagem rápida e econômica, funcionando como base para liberação de laudos de constatação provisória. Ao passo que testes analíticos mais caros e demorados (como o GC–MS, por exemplo) possam ser priorizados em análises definitivas de drogas de abuso, que segundo o artigo 159 do Código de Processo Penal (CPP), deverá ser realizado em até dez dias e prorrogado em casos especiais. Sendo assim, os testes de cores devem constituir-se em parte integrantes do processamento de rotina de substâncias suspeitas de tratar-se de uma droga ilícita [1].

A multiplicidade de Novas Substâncias Psicoativas (NSP) desafia a eficácia do atual aparato técnico para identificação preliminar dessas drogas. O desenvolvimento de novos métodos analíticos para detecção e identificação de NSP é fundamental para análise de química forense [4]. Nesse contexto, os canabinoides sintéticos representam um dos grupos de drogas sintéticas mais numerosas e incidentes da atualidade [5]. E ao contrário da maconha clássica e tradicional, a “maconha sintética” não é presumivelmente identificada utilizando o teste colorimétrico *Duquenoís-Levine* e *Fast Blue B*. Além disso, o teste de cor de *van Urk*, utilizado para identificação presuntiva de indol, é negativo para as principais classes de canabinoides sintéticos [2]. No que tange a realidade atual do Laboratório Forense da Polícia Científica do Estado do Pará, especificamente a coordenadoria regional II em Marabá, não se dispõe de reagentes colorimétricos para identificação preliminar de canabinoides sintéticos.

Diante de situações que demandam a realização de exame de constatação provisória dessas drogas, deve valer-se de metodologias e equipamentos de constatação definitiva, mais onerosos e que requerem maior complexidade técnica, por isso, tendem a ser mais demorados [6].

Sendo assim, é de extrema relevância encontrar substâncias com potencial de serem utilizadas no desenvolvimento de um protocolo de teste simples, rápido, barato e seletivo para triagem colorimétrica das principais classes de canabinoides sintéticos.

Diante disso, os métodos de Triagem Virtual inovam na descoberta de novos compostos com atividade específica, avaliando *in silico* grandes bibliotecas estruturais contra um alvo químico específico, favorecendo assim a redução de esforços financeiros, infraestrutura e tempo envolvidos no processo de descoberta de novas quimioestruturas [7]. Esses métodos aplicam etapas sequenciais e hierárquicas que visam filtrar e selecionar compostos com propriedades físico-químicas de interesse, descartando aqueles que não se encaixam nas características desejáveis [8].

Desse modo, o presente estudo propõe aplicar a Triagem Virtual baseada em ligantes na seleção de compostos químicos que possam apresentar potencial atividade colorimétrica para identificação de canabinoides sintéticos.

2. MATERIAIS E MÉTODOS

2.1 Caracterização metodológica

Foi realizado um estudo teórico de química computacional, objetivando encontrar compostos com similaridade molecular às substâncias de referência e que possuam potencial aplicação para indicação colorimétrica da maioria de canabinoides sintéticos.

A Triagem Virtual é a ferramenta computacional utilizada para seleção de compostos de interesse em biblioteca de compostos químicos, sendo a Triagem Virtual baseada em ligantes, a abordagem aplicada para o estudo.

2.2 Etapa 1

Preliminarmente realizou-se uma pesquisa da literatura científica e acadêmica para seleção de artigos científicos que apresentassem compostos químicos com atividade colorimétrica para canabinoides sintéticos validada experimentalmente, permitindo, dessa forma, serem utilizados como referência ou molde para a Triagem Virtual baseada em ligantes. Para isso foram exploradas as plataformas Google acadêmico, ResearchGate, Scielo e PubMed, usando filtros que consideravam apenas publicações nos idiomas Português, Inglês e Espanhol, dentro do intervalo de 2009 a 2023. A

pesquisa foi estabelecida a partir de 2009, pois foi o ano em que foi publicado na literatura internacional, a primeira identificação de canabinoides sintéticos em ervas [9].

Partindo desse princípio, foram selecionadas as substâncias 4-(dimetilamino)benzaldeído (p-DMAB, reação de Ehrlich) e 2,4-dinitrofenil-hidrazina (2,4-DNPH, reagente de Brady) [2,10].

2.3 Etapa 2

Desenvolveu-se o processo de filtragem molecular dentro da base de dados de compostos químicos da *PubChem* [11], utilizando descritores bidimensionais (2D) ou características moleculares específicas (*fingerprint molecular*) e empregando o Coeficiente de Tanimoto, como medida de similaridade. Além disso, foram utilizados parâmetros físico-químicos de comparação como massa molecular, clogP, contagem de doadores de hidrogênio, contagem de receptores de hidrogênio, contagem de ligações rotacionáveis, área polarizada, dentre outros (Tabela 1).

2.4 Etapa 3

A relação de substâncias triadas e selecionadas, na etapa anterior, foram salvas em formato de arquivo SDF, e transferidas para *software Osires DataWarrior* [12]. Esta ferramenta computacional possibilitou a realização de processo de filtragem complementar, com base em duas metodologias de comparação por similaridade molecular: a *Principal Component Analysis* (PCA) e a *Cluster Analysis*, empregando o algoritmo *Self Organizing Maps* (SOM) com o *fingerprint* molecular FragFP.

Ademais, realizou-se a comparação de substâncias analisando subestruturas, derivadas computacionalmente na forma de andaimes moleculares, por remoção de átomos de cadeias laterais. Dar-se o nome de Análise de *Scaffolds*, que, para o presente estudo, baseou-se no sistema de anel fechado e heteroátomo [13].

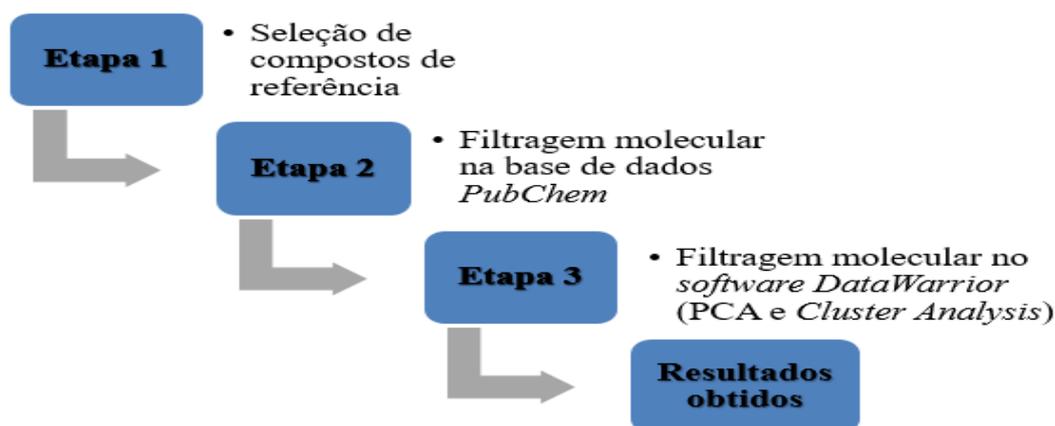


Figura 1. Fluxograma da metodologia utilizada

3. RESULTADOS E DISCUSSÕES

3.1. Seleção de moléculas de referência

Uma maneira de identificar testes de cor adequados para novas classes de compostos é considerar como e por que os testes de cor estabelecidos funcionam ou não do seu ponto de vista estrutural, partindo-se da premissa de que moléculas com estruturas semelhantes apresentam propriedades físico-químicas e comportamentos semelhantes [14].

Nesse contexto, para o presente estudo, duas substâncias foram utilizadas como referência ou molde para Triagem Virtual baseada em ligantes. A primeira substância foi o 4(dimetilamino)benzaldeído (p-DMAB, reação de Ehrlich). Na referência [10] foi objeto de estudo para o desenvolvimento de ensaio colorimétrico simples

para identificação do grupo aminoalquilindol, um dos maiores grupos de canabimiméticos sintéticos. A Figura 2, caracteriza a reação de Ehrlich, uma reação de substituição eletrofílica de indóis com compostos carbonílicos, catalisada por sílica e ácido sulfúrico, que resulta na formação de bis(indolil)alcanos, implicando em mudança de cor de amarelo para roxo/azul, quando positivo [15].

O método proposto demonstra tratar-se de um procedimento colorimétrico relativamente seletivo e sensível para a determinação de canabimiméticos sintéticos contendo aminoalquilindol em análise de amostra única [10].

O 2,4-dinitrofenilhidrazina (2,4-DNPH, reagente de Brady) é a segunda substância de referência. Na referência [2] foi investigado seu uso como um reagente de cor alternativo para identificação presuntiva de

canabinoides sintéticos, tendo como alvo da reação, o grupo químico carbonila, ao invés do grupo indol. Consiste em um mecanismo de condensação (Figura 3), caracterizado por uma reação nucleofílica de adição-eliminação em que o 2,4-DNPH, em um primeiro momento, adiciona-se a ligação dupla da carbonila (presente no canabinoide sintético) para dar origem a um composto intermediário que perde uma molécula de água (H₂O). O precipitado da reação varia entre uma cor amarela, laranja ou vermelho [16].

Os autores da referência [17] conduziram um estudo que confirma a tese de que o grupo carbonila, presente em aldeídos e cetonas, seria uma característica estrutural que oferecia um potencial reativo para detecção específica deste grupo funcional, utilizando o DNPH como agente da reação. Os canabimiméticos sintéticos do grupo dos naftoilindol, benzoilindole, fenilacetilindole e cicloalcoilindole contêm um grupo funcional cetona diretamente ligado a um anel aromático, tonando-se

potenciais substâncias passíveis de identificação com reagente de cor que sinaliza a presença de carbonilas.

A imensa maioria das reações colorimétricas de interesse forense ocorrem entre compostos eletrofílicos e nucleofílicos. A mistura dos dois pode permitir que o componente de droga, que em geral é “rico em elétrons” doe densidade eletrônica ao componente de teste “pobre em elétrons”. Se a energia associada ao processo de transferência de elétrons corresponder à região visível do espectro eletromagnético, a transferência pode se manifestar visualmente como uma mudança de cor (resultado positivo para o teste) [18].

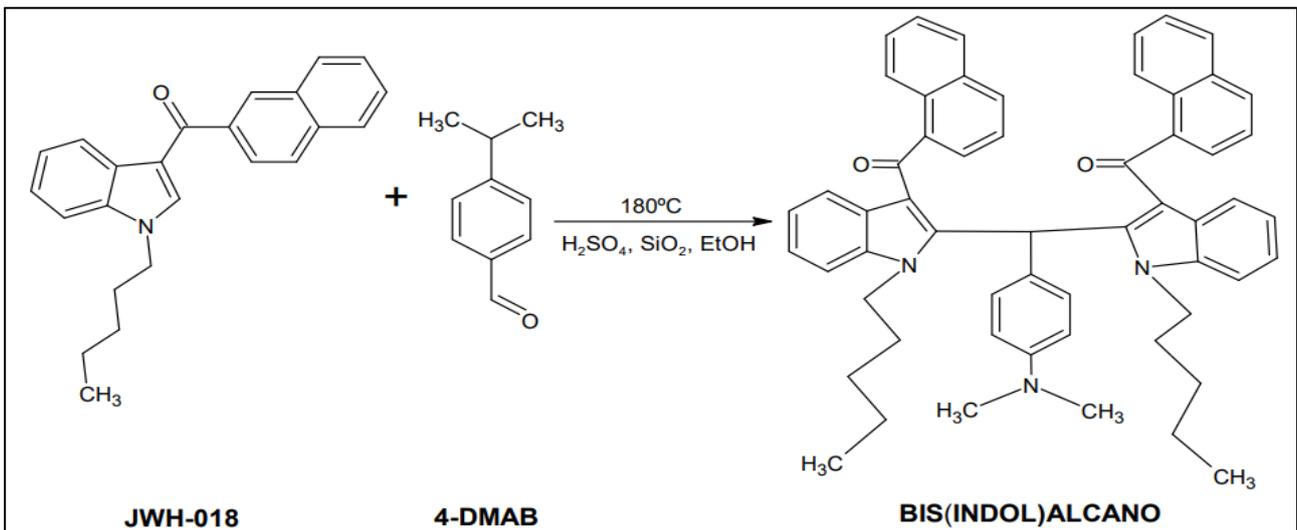


Figura 2. Reação de Ehrlich.

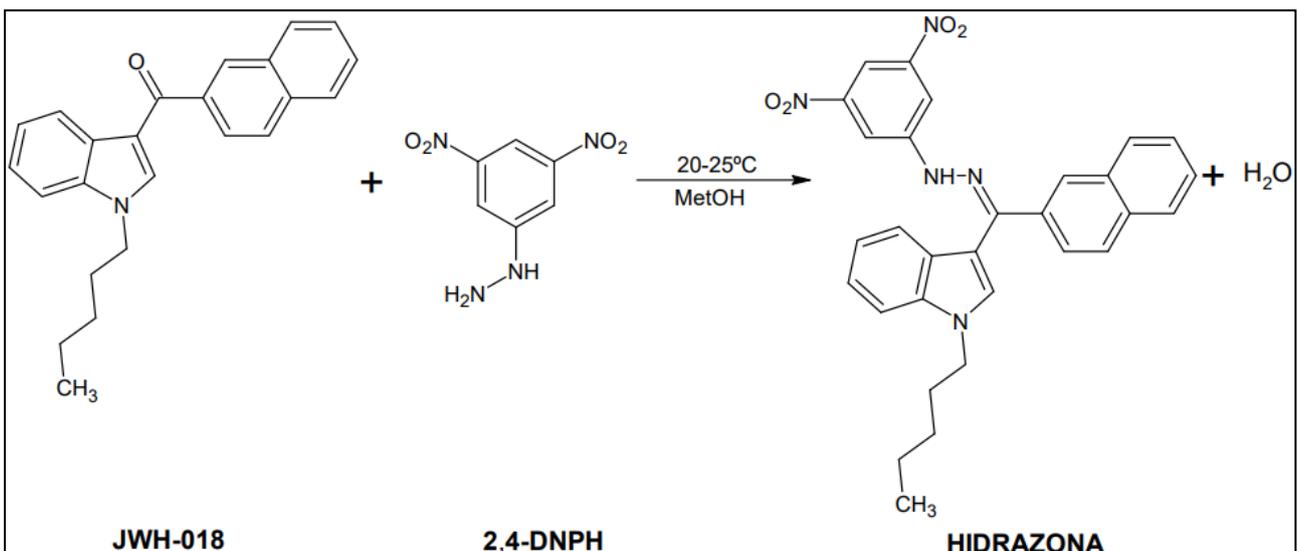


Figura 3. Reação de Brady.

3.2. Biblioteca virtual de compostos químicos

A utilização de uma biblioteca virtual de compostos químicos é uma etapa crucial para execução de procedimentos de Triagem Virtual, pois fornece compostos com o maior grau possível de diversidade molecular, sejam eles já conhecidos e consagrados por suas atividades biológicas e terapêuticas, sejam compostos já sintetizados e disponíveis para ensaios biológicos, moléculas hipotéticas (desenhadas em computador antes da síntese) ou até produtos de origem natural. A qualidade da base de dados utilizada é de vital importância, pois é a fonte de onde serão selecionados os compostos promissores para futuros ensaios experimentais [19].

Para este estudo foi utilizado o *PubChem* [11], como o banco de dados químico para obtenção de compostos com similaridade molecular às estruturas utilizadas como referência. Segundo a referência [20], o *Pubchem* contém um dos maiores acervos de informações químicas

disponíveis publicamente, organizado em três bases de dados interligadas: *Substance*, *Compound* e *BioAssay*. Em Li *et al.* [21] reitera-se que o *PubChem* reduz as barreiras para que pesquisadores avancem no desenvolvimento de ferramentas químicas que possam modelar processos biológicos, apresentar potenciais candidatos a fármacos para tratamentos de doenças e indicar novas ferramentas químicas com as mais diversas aplicabilidades.

Medir a similaridade química entre duas substâncias significa comparar com um padrão de características pré-definidas [19]. Neste sentido, foi realizada uma pesquisa de estruturas com similaridade química aos compostos de referência, considerando uma similaridade bidimensional (2D), baseada em coeficiente de Tanimoto. Importa salientar que para filtragem molecular nessa base de dados foram empregadas propriedades químicas específicas das substâncias de referências (Tabela 1), funcionando como filtros moleculares quimioinformáticos [11].

Tabela 1. Principais propriedade físico-químicas das substâncias utilizadas como referência para filtragem molecular no *PubChem*.

PROPRIEDADES FÍSICO-QUÍMICAS	SUBSTÂNCIAS REFERÊNCIA	
	4-DMAB	2,4-DNPH
FÓRMULA MOLECULAR	C ₉ H ₁₁ NO	C ₆ H ₆ N ₄ O ₄
MASSA MOLECULAR g/mol	149,19	198,14
CLOGP	1,8	1,5
CONTAGEM DE DOADORES DE H ⁺	0	2
CONTAGEM DE RECEPTOR DE H ⁺	2	6
CONTAGEM DE LIGAÇÕES ROTACIONÁVEIS	2	1
MASSA EXATA	149	198
MASSA MONOISOTÓPICA	149	198
ÁREA DE SUPERFÍCIE POLAR TOPOLÓGICA	20,03	130
CONTAGEM DE ÁTOMOS PESADOS	11	14
CONTAGEM DE ÁTOMOS ISOTÓPICOS	0	0

Tabela 2. Substâncias com similaridade química ao 4-DMAB.

COMPOSTOS	NOME (IUPAC)	MM g/mol	NHBD	NHBA	Nrotb	NAN	clogP
Composto 1	2-(dimethylamino)benzaldehyde	149	0	2	2	1	1,5
Composto 2	4-(dimethylamino)(14C)benzaldehyde	151	0	2	2	1	2
Composto 3	3-(dimethylamino)benzaldehyde	149	0	2	2	1	1,5
Composto 4	[4-(dimethylamino)phenyl]methanone	148	0	3	4	1	3,9
Composto 5	4-(aziriden-1-yl) benzaldehyde	147	0	2	2	1	1,3
Composto 6	deutério-[4-(dimethylamino)phenyl]methanone	150	0	2	2	1	1,8
Composto 7	2-deutério-4-(dimethylamino)benzaldehyde	150	0	2	2	1	1,8
Composto 8	4-[bis(trideuterio methyl)amino]-2,3,5,6-tetradeuterio-benzaldehyde	159	0	2	2	1	1,8
Composto 9	(4-formylphenyl)-methyl-azanide;rubidium(1+)	218	1	2	2	1	1,2
Composto 10	deuterio-[4-(dimethylamino)phenyl](113c)methanone	151	0	2	2	1	1,8
Composto 11	4-(dimethylamino)(18O)benzaldehyde	151	0	2	2	1	1,8
Composto 12	3,5-dideuterio-4-(dimethylamino)benzaldehyde	151	0	2	2	1	1,8
Composto 13	4-(dimethylamino)(13C)benzaldehyde	150	0	2	2	1	1,8

Após o processo de filtragem molecular dentro da biblioteca virtual procedente do *PubChem*, foi possível obter dois conjuntos de compostos com estruturas similares às substâncias referências 4-DMAB e 2,4-DNPH, respectivamente (Tabelas 2 e 3).

3.3. Filtragem molecular no Osires DataWarrior

De acordo com a referência [22], para explorar visualmente os dois conjuntos de compostos (Tabelas 2 e 3), o *software Osires DataWarrior* utiliza uma abordagem direta, em que as moléculas são representadas individualmente, incorporadas em um plano bidimensional e tridimensional (2D e 3D,

respectivamente), com base em propriedades selecionadas e empregando técnicas de redução de dimensionalidade ou escalonamento multidimensional. Nessa abordagem de visualização direta, as moléculas são desenhadas como entidades individuais, não existindo um modo universal de mapear moléculas em coordenadas num espaço euclidiano. Assim, começando com um conjunto de moléculas a serem visualizadas, o processo de cálculo de suas coordenadas é geralmente bifásico, sendo primeiro criado um espaço de propriedades que serão projetadas para um espaço euclidiano bidimensional ou tridimensional.

Tabela 3. Substâncias com similaridade química ao 2,4-DNPH.

COMPOSTOS	NOME (IUPAC)	MM g/mol	NHBD	NHBA	Nrotb	NAN	clogP
Composto 1	(2,4-dinitro(1,2,3,4,5,6-13C6)cyclohexa-1,3,5-trien-1-yl)hydrazine	204	2	6	1	4	1,5
Composto 2	4-diazenyl-6-nitro-bezene-1,3-diamine	181	3	7	1	5	1,5
Composto 3	(3,5-dinitrophenyl)hydrazine	198	2	6	1	4	0,8
Composto 4	(2,3-dinitrophenyl)hydrazine	396	4	12	2	8	1,4
Composto 5	(3,4-dinitrophenyl)hydrazine	198	2	6	1	4	0,8
Composto 6	(2,6-dinitrophenyl)hydrazine	198	2	6	1	4	1,0
Composto 7	(2,5-dinitrophenyl)hydrazine	198	2	6	1	4	1,5
Composto 8	(2,4-dinitrophenyl)diazene	356	0	6	2	4	5,2
Composto 9	[2,4-bis[oxidanylidene(oxido)(15N)(15N)ammonio]phenyl] hydrazine	204	2	8	3	4	1,5
Composto 10	[2,4-bis[oxidanylidene(oxido)ammonio]phenyl] hydrazine	202	2	8	3	4	1,5
Composto 11	(2,4-dinitrophenyl)(15N)hydrazine	200	2	6	1	4	1,5
Composto 12	[2,4-bis[oxido(oxo)(15N)ammonio]phenyl] hydrazine	200	2	8	3	4	1,5
Composto 13	(2,4-dinitroanilino)ammonium;	199	2	8	3	4	1,5
Composto 14	(2,4-dinitrophenyl)hydrazine;hydrochloride	234	3	6	1	4	1,5
Composto 15	(2,3-dinitrophenyl)hydrazine;hydrochloride	234	3	6	1	4	1,5
Composto 16	(2,4-dinitroanilino)ammonium;chloride	234	2	8	3	4	1,5
Composto 17	amino-(2,4-dinitrophenyl)ammonium; hydrohoide	235	2	8	3	4	1,5
Composto 18	N-hydroxy-2,4-dinitro-benzeniamine oxide	215	2	9	3	3	1,2
Composto 19	1,1-dideuterio-2-(2,4-dinitrophenyl)hydrazine	200	2	8	3	4	1,5
Composto 20	(2,3,5-trideuterio-4,6-dinitrophenyl)hydrazine	201	2	8	3	4	1,5

O espaço de propriedades é obtido escolhendo um conjunto de propriedades moleculares que supostamente

guiará a visualização. A partir dessas propriedades, obtemos coordenadas em um espaço euclidiano de alta dimensão (por exemplo, escolhendo propriedades numéricas e usando os valores das propriedades como coordenadas) ou definimos uma métrica baseada na similaridade das moléculas, produzindo um espaço métrico. Uma vez definido um espaço de propriedades, um método de projeção adequado é usado para calcular as coordenadas desejadas. Os dados como massa molecular (MM), número de doadores de hidrogênio (NHBD), número de aceptores de hidrogênio (NHBA), número de ligações rotacionáveis (Nrotb), número de átomos de nitrogênio (NAN) e clogP (coeficiente de partição), são amplamente utilizadas na maioria dos métodos de avaliação *in silico* (Tabelas 2 e 3) [23-25].

Para Russel [26], a massa molecular (MM) é um importante descritor físico-químico que indica o tamanho da cadeia e a massa total de uma molécula, constituindo-se da soma de massas atômicas de todos os átomos presentes na fórmula molecular da substância. Relacionando essa propriedade com os compostos similares ao 4-DMAB, tem-se uma massa molecular média de 156 g/mol, apenas 5% superior ao MM da molécula referência (149 g/mol), demonstrando tratar-se de compostos que apresentam valores de massas moleculares próximos entre si. Em contrapartida, o composto 9 foi o que apresentou variação dissonante de MM em relação à média e a substância referência, aproximadamente, 28% e 31% superior, respectivamente. Já com relação aos compostos similares ao 2,4-DNPH, a massa molecular média é de 225 g/mol, sendo superior em 12% ao MM da molécula referência (198 g/mol). Os compostos 4 e 8 foram os que demonstraram maior variação com relação à massa molecular da média e da molécula referência, respectivamente, 44%, 50% e 29%, 45%.

O número de doadores de hidrogênio (NHBD), também conhecida como número de ligações de hidrogênio, é uma medida que descreve a capacidade de uma molécula doar átomos de hidrogênio para formar ligações de hidrogênio com outras moléculas. Isso ocorre quando um átomo de hidrogênio, ligado a um átomo eletronegativo como o oxigênio, nitrogênio ou flúor, é atraído por forças intermoleculares, formando uma interação intermolecular de alta polaridade entre as moléculas [27]. Considerando a aplicação que se deseja, a característica é de extrema relevância pois o NHBD pode influenciar a capacidade de uma substância, com suposta ação colorimétrica, interagir com um grupo específico de droga de abuso através de ligações de hidrogênio. No que se refere ao grupo de compostos similares ao 4-DMAB, apenas o composto 9 possui, na sua estrutura, grupo de átomos doadores de hidrogênio, ao passo que do grupo de compostos similares ao 2,4-DNPH, apenas o composto 8 não possui doadores de hidrogênio.

O número de aceptores de hidrogênio (NHBA) é uma medida da capacidade de uma molécula de aceitar átomos de hidrogênio em ligações de hidrogênio com outras moléculas. Os aceptores de hidrogênio são átomos eletronegativos, como oxigênio ou nitrogênio, que possuem pares de elétrons livres capazes de formar ligações de hidrogênio com átomos de hidrogênio de outras moléculas [27-28]. Vale acrescentar que todos os compostos similares ao 4-DMAB possuem 2 ou mais aceptores de hidrogênio, ao passo que para aqueles com similaridade ao 2,4-DNPH, apresentam 6 ou mais aceptores de hidrogênio, destaque para o composto 4, com 12.

O número de ligações rotacionáveis (Nrotb), é uma característica que representa bem a flexibilidade molecular, e está intimamente relacionado à afinidade do alvo e à força da interação intermolecular, ao passo que a diminuição de Nrotb torna a molécula mais rígida e, portanto, restringe sua liberdade conformacional, enquanto um grande aumento de Nrotb pode aumentar a complexidade molecular e, assim, influenciar sua atividade biológica, podendo ligar-se a distintos alvos moleculares [29-30]. Todos os compostos similares ao 4-DMAB possuem 2 ou mais ligações rotacionáveis, enquanto os compostos com similaridades ao 2,4-DNPH, 1 ou mais ligações.

O número de átomos de nitrogênio (NAN) em uma reação química é uma variável crítica que pode influenciar a estequiometria, a reatividade, o mecanismo de reação e as propriedades finais dos produtos, tornando-se característica essencial para a compreensão completa e a execução precisa de reações químicas, pois pode servir como constituinte de proteção ou direcionador em síntese orgânica. O comportamento dos átomos de nitrogênio em reações químicas é multifacetado e depende do ambiente químico e do tipo de reação, podendo atuar como nucleófilos, bases, agentes redox, e desempenhar papéis cruciais em várias transformações químicas. Além disso, nitrogênios adicionais podem introduzir diferentes grupos funcionais (como aminas, amidas, nitrilos) que afetam a reatividade, solubilidade, polaridade, e outras características [31]. No tocante aos compostos similares ao 4-DMAB e ao 2,4-DNPH, todos apresentam átomos de nitrogênio.

O clogP é uma abreviação para "logP calculado", em que "P" representa "partição". É uma medida que descreve a lipofilia ou hidrofobicidade de uma molécula, representando a razão entre a concentração de uma molécula em uma fase orgânica e a concentração na fase aquosa, em condições específicas. Quanto maior o valor de logP, mais lipofílica é a molécula, indicando uma maior afinidade pela fase orgânica em relação à fase aquosa. Muitas vezes é determinado por meio de métodos computacionais, como modelos de previsão baseados em estrutura molecular, que estimam o logP de uma molécula

com base em sua estrutura química. Esses modelos podem considerar vários parâmetros, como a distribuição de grupos funcionais na molécula e a polaridade relativa de diferentes regiões da molécula [27]. Para os compostos com similaridade ao 4-DMAB, os valores de logP variaram de 1,2 a 3,9, sendo assim, considerados, em sua maioria, moderadamente lipofílicos (logP entre 1 e 3). Ao passo que, os compostos com similaridade ao 2,4-DNPH obedecem ao intervalo entre 0,8 a 5,2, indicando a variação da propriedade com relação a este grupo de compostos, pois apresentam composto mais lipofílicos do que hidrofílicos (logP > 0), moderadamente lipofílicos (logP entre 1 e 3) e altamente lipofílicos (logP > 3).

Para Velkoborsky e Hoksza [32], a análise de *Scaffolds* compreende à investigação e à avaliação dos componentes estruturais fundamentais compartilhados entre diferentes compostos químicos em uma série de moléculas relacionadas. Assim, um *Scaffold*, ou esqueleto ou andaime molecular, é a estrutura central de uma molécula que serve como base para compreensão das relações entre estrutura química e atividades. Nesse contexto, a referência [33] complementa que além da utilização dessa abordagem ser importante para entender as relações estrutura-atividade (SAR, do inglês *Structure-Activity Relationship*), ela orienta no processo de descoberta de novas estruturas químicas, com atividades diversas, por envolver a identificação de grupos funcionais, anéis e outras características estruturais comuns entre as moléculas de uma série, bem como a determinação de como essas características contribuem para suas atividades.

A análise de *Scaffolds* dos compostos relacionados nas Tabelas 2 e 3, permitem inferir que apresentam similaridade acentuada no seu esqueleto molecular pois possuem a mesma estrutura central ou o mesmo padrão básico de ligações entre os átomos que compõem a molécula. Posto isto, quando os compostos compartilham do mesmo esqueleto molecular, propende a apresentar propriedades químicas e biológicas semelhantes. No entanto, pequenas modificações na estrutura molecular, como a adição de grupos funcionais ou a alteração da conectividade dos átomos, podem resultar em diferenças significativas nas propriedades físico-químicas e na atividade biológica dos compostos.

Dentre as funcionalidades do *DataWarrior*, a *Principal Component Analysis* (PCA) e a *Cluster Analysis*, que utiliza *Self Organizing Maps*, foram selecionadas e aplicadas como base de comparação de similaridade molecular para o presente estudo, corroborando com a referência [34], que enfatiza a importância dessas técnicas como métricas de similaridade e distância, aplicadas para comparar impressões digitais de compostos químicos.

Para definir a comparação entre duas moléculas foi necessário empregar (tanto para análise de PCA como

para a *Cluster Analysis*) um padrão de medida ou coeficiente de similaridade, o qual quantificará o grau de semelhança entre as entidades químicas avaliadas. O coeficiente ou índice escolhido para este estudo foi o de Tanimoto (S_{TAN}) que fornece valores de comparação na faixa de 0 (zero) a 1 (um). O valor de 0 (zero) significa que não há similaridade entre as moléculas comparadas, já o valor de 1 (um) consiste em 100% de similaridade. O valor de *Cutoff* (ou ponto de corte) utilizado foi de 0,7 ou superior, como um limiar para indicar que duas estruturas são consideradas suficientemente semelhantes. A referência [35] ratifica que para muitas análises de similaridade, especialmente na análise de dados de expressão gênica, bioinformática e quimiometria, um *Cutoff* de 0,7 é convenientemente utilizado como uma medida adequada para determinar uma forte similaridade entre amostras ou variáveis em análise. Em outras palavras, isso significa dizer que estruturas com um coeficiente de Tanimoto de 0,7 ou superior são consideradas como tendo uma sobreposição significativa em suas características e, portanto, possuindo uma similaridade relevante [36-37]. Através do coeficiente de Tanimoto, calcula-se a razão dos números de características comuns entre as moléculas, pelo número total de características presentes em ambas, subtraído pelo número de características comuns [19].

Segundo a referência [38], o coeficiente de Tanimoto tem sido o mais utilizado para esse fim, pois demonstra ser um padrão de confiança, considerando os resultados da sua utilização nos mais diversos experimentos. As referências [34,39] ratificam o coeficiente de Tanimoto, como o mais amplamente aplicado em similaridade baseada em impressões digitais, com resultados equivalentes a outras métricas de similaridade aplicadas para comparar duas moléculas, como Soergel, Dice e Cosine, enquanto as medidas de similaridade derivadas das distâncias euclidiana e de Manhattan são descritas como insatisfatórias.

A PCA é uma técnica estatística multivariada utilizada para simplificar a complexidade de conjuntos de variáveis, identificar padrões e estruturas subjacentes, e reduzir a dimensionalidade dos dados. Dessa forma, é capaz de identificar semelhanças e diferenças entre compostos químicos submetidos a análise, além de definir o comportamento destas em relação a um parâmetro de medida. Nesse sentido, o cálculo de PCA permite condensar as informações contidas em várias variáveis originais correlacionadas em um conjunto menor de variáveis estatísticas linearmente não correlacionadas, chamadas componentes principais [40-41].

No estudo foi utilizado três componentes principais (PC1, PC2 e PC3) para oferecer uma visão clara e simplificada da estrutura multidimensional dos dados de ligantes. O PC1 é a direção que captura a maior parte da variação total presentes no conjunto de dados originais,

representando as características mais dominantes que diferenciam os ligantes e sendo o eixo em que os dados projetados têm a maior dispersão. Já o PC2 é a direção que captura a segunda maior parte da variação total, sob a restrição de ser ortogonal ao PC1 e com cargas diferentes. E o PC3 captura a terceira maior parte de variação total, sendo ortogonal tanto para PC1 quanto a PC2 e com cargas específicas [42].

Outro conceito importante é a matriz de cargas (ou *loadings*), que demonstra a correlação entre as variáveis originais e os componentes principais PC1, PC2 e PC3. Cada elemento da matriz representa a contribuição de uma variável original para um componente principal, apresentando valores de carga que variam de -1 a 1. Os valores próximos de 1 ou -1 indicam uma forte correlação positiva ou negativa, respectivamente, entre a variável original e o componente principal, enquanto valores próximos de 0 indicam pouca ou nenhuma correlação [40,42].

Na análise de PCA, a dimensão de representação gráfica dos dados ocorreu na forma 2D e 3D. Os descritores 2D consideram tanto as propriedades dos átomos quanto as suas conectividades. Já os descritores 3D possuem uma maior complexidade, pois também consideram as conformações das moléculas, sendo aplicados para representar distâncias inter-atômicas, farmacofóricos 3D, superfícies moleculares, campos eletrostáticos e outros [43]. Estudos demonstraram que o coeficiente de similaridade 2D negligencia algumas características estruturais/funcionais importantes na identificação do composto alvo [44].

Os Gráficos 1 e 2 caracterizam uma representação em gráfico de dispersão desenvolvida a partir de descritores 2D ou características moleculares específicas (“*fingerprints*”) derivadas das estruturas de compostos de referência e comparados com os correspondentes

descritores das moléculas similares provenientes da base de dados.

Os Gráficos 3 e 4 são representações 3D em gráfico de dispersão. Como dito, os métodos baseados em descritores 3D são mais elaborados, sendo a maioria dependentes do alinhamento de conformações moleculares. Talvez o método mais amplamente empregado exigindo representações 3D, é aquele que explora o conceito de similaridade aos grupos da molécula essenciais à atividade (farmacóforo) [45].

É relevante salientar, que o programa utilizado visualiza o conjunto de compostos na forma de descritores moleculares que podem ser codificados em vetores ou *fingerprints*, na grande maioria binários. Para o presente estudo foi selecionado um descritor baseado em *fingerprints* molecular, o descritor FragFP, que se baseia na fragmentação da molécula em fragmentos menores e na geração de *fingerprints* a partir desses fragmentos. Trata-se de representações binárias (bits) de uma estrutura química em que 1 (presente) denota a existência de uma determinada característica molecular e 0 (ausente) denota a inexistência. A combinação dessas informações forma o *fingerprint* molecular, o que ofereceu uma representação compacta e informativa das moléculas, permitindo uma análise rápida e eficiente de grandes conjuntos de dados moleculares [46-47].

Gráfico 1. Análise de PCA, utilizando descritores 2D e considerando o 4-DMAB como substância de referência.

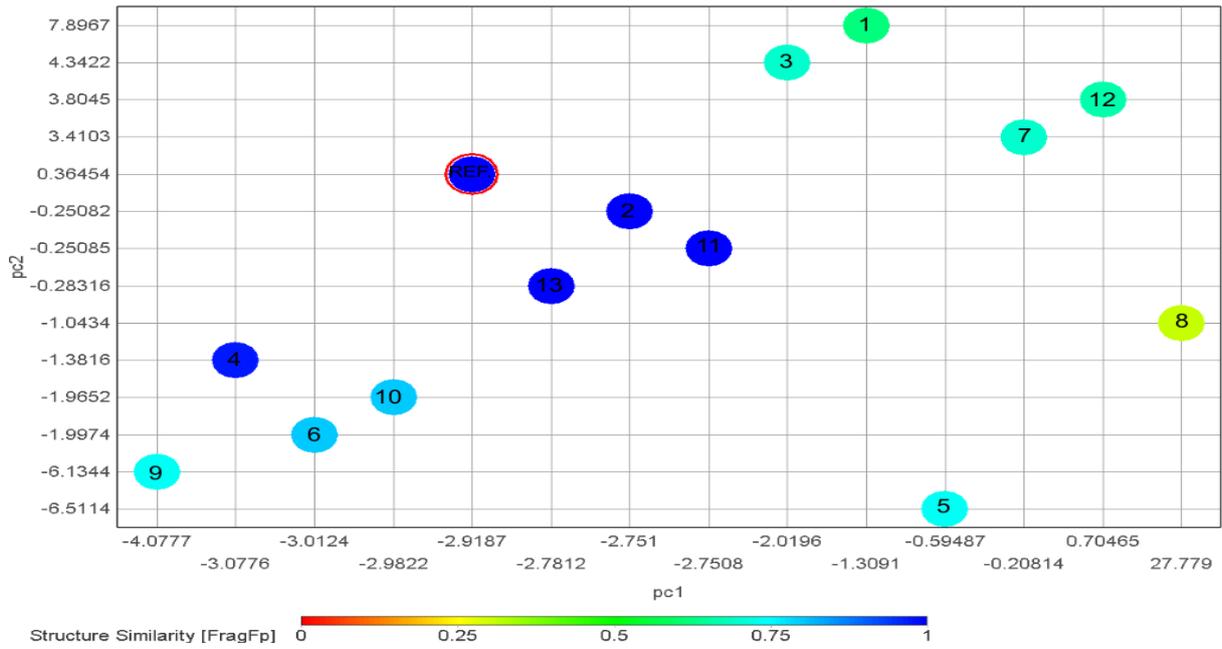


Gráfico 2. Análise de PCA, utilizando descritores 2D e considerando o 2,4-DNPH como substância de referência.

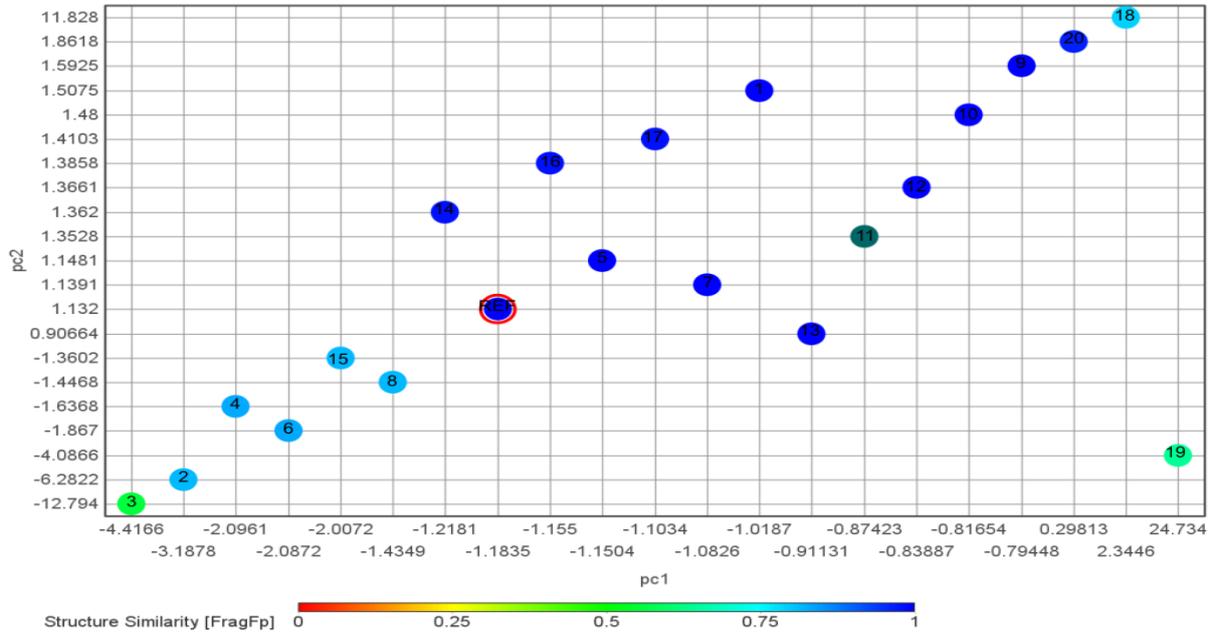


Gráfico 3. Análise de PCA, utilizando descritores 3D e considerando o 4-DMAB como substância de referência.

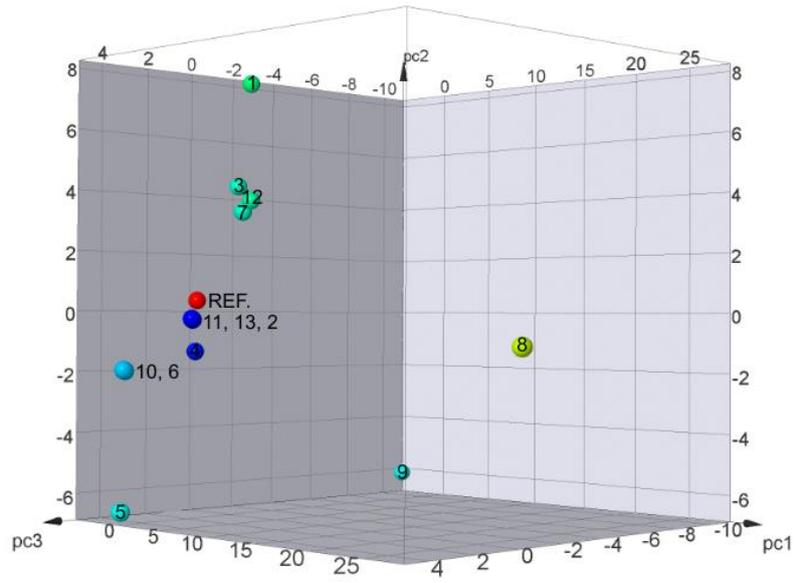
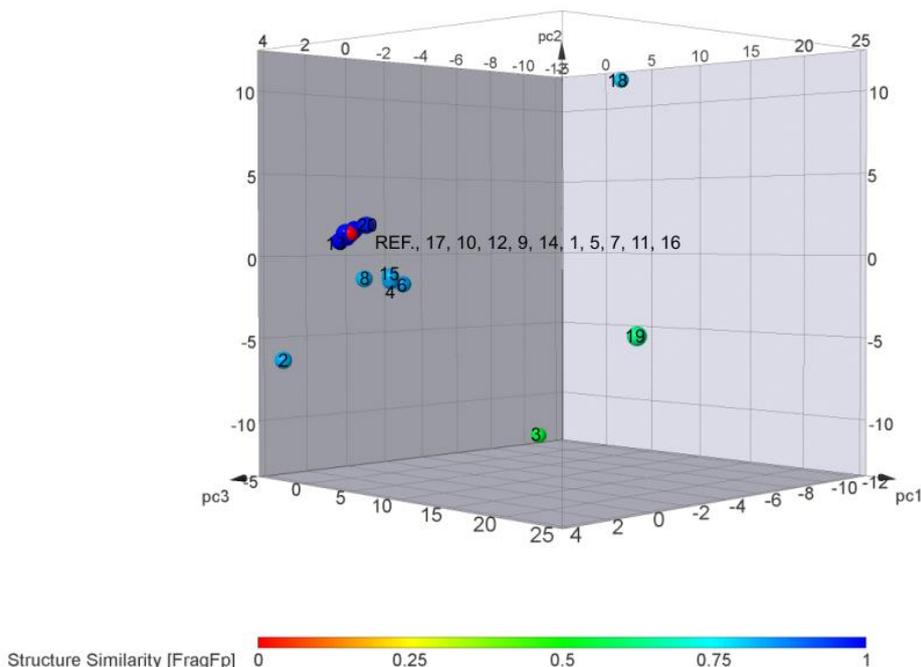


Gráfico 4. Análise de PCA, utilizando descritores 3D e considerando o 2,4-DNPH como substância de referência.

Os Gráficos 5 e 6 representam a *Cluster Analysis*, uma outra abordagem utilizada pelo *DataWarrior* para comparação de similaridade molecular com os compostos de referência, baseado em um método hierárquico divisivo, utilizando a similaridade baseada em impressões digitais moleculares, para o agrupamento de compostos. Dessa forma, foi criada uma hierarquia baseada em características estruturais das moléculas, ou seja, as moléculas que partilham características estruturais comuns num determinado nível são agrupadas [48]. A análise de agrupamento é, portanto, uma técnica de

classificação multivariada que pode ser utilizada quando se deseja explorar as similaridades entre casos, indivíduos, objetos ou entre variáveis, definindo-os em grupos. Assim, o objetivo principal da aplicação da *Cluster Analysis* para esse estudo é dividir as estruturas em grupos ou *clusters*, de modo que os compostos químicos dentro de cada grupo sejam mais semelhantes entre si do que com aqueles em outros grupos [49].

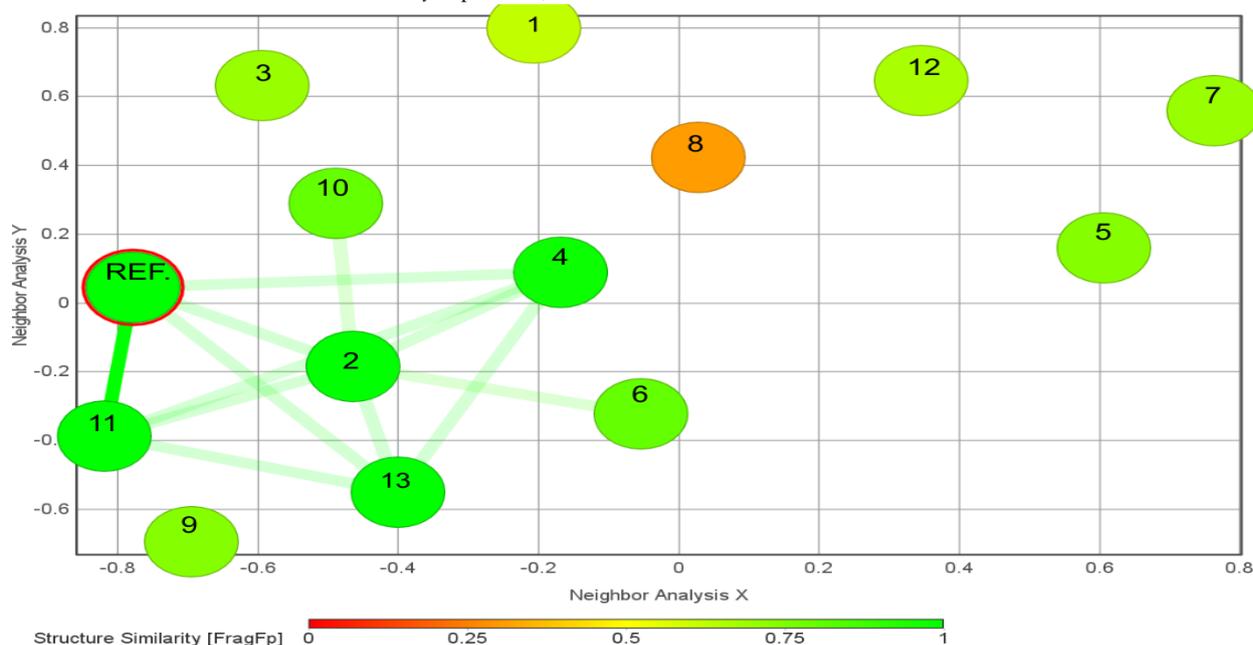
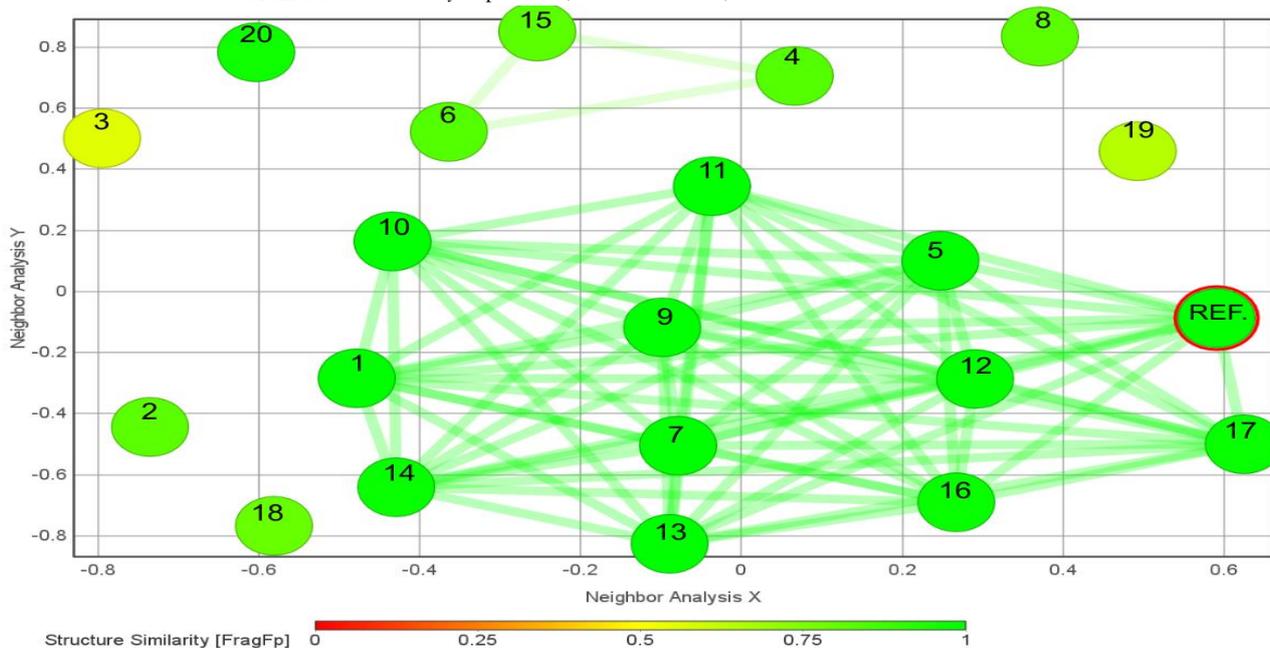
Gráfico 5. *Cluster Analysis* por SOM, considerando o 4-DMAB como substância de referência.

Gráfico 6. Cluster Analysis por SOM, considerando o 2,4-DNPH como substância de referência.

Considerando o valor de *Cutoff* de 0,7 (S_{TAN}), como limiar de similaridade predefinido para comparação com as substâncias de referência ou molde, apresentamos nas Tabelas 4 e 5, a relação de compostos que possuem potencial atividade colorimétrica para identificação preliminar de canabinoides sintéticos, partindo-se da premissa de que moléculas com estruturas semelhantes apresentam propriedades físico-químicas e comportamento semelhantes. Após o uso de ferramenta computacional, com aplicação de métodos de comparação por similaridade (PCA e *Cluster Analysis*), foi possível determinar que dos compostos apresentados nos Quadros 1 e 2, apenas os compostos 1, 7, 8 e 12, não possuem similaridade aceitável com o 4-DMAB, e os compostos 3 e 19 não possuem similaridade aceitável com o 2,4-DNPH.

Tabela 4. Valores do Coeficiente de Tanimoto para os compostos com similaridade molecular ao 4-DMAB.

COMPOSTOS	PubChem CID	S_{TAN}
Composto 1	34289	0,62
Composto 2	101720972	1
Composto 3	34288	0,70
Composto 4	85602031	0,97
Composto 5	45083593	0,74
Composto 6	101946788	0,80
Composto 7	53680852	0,69
Composto 8	53680853	0,30
Composto 9	57673085	0,74
Composto 10	101946789	0,80
Composto 11	101946791	1
Composto 12	101946792	0,67
Composto 13	101946790	1

Tabela 5. Valores do Coeficiente de Tanimoto para os compostos com similaridade molecular ao 2,4-DNPH.

COMPOSTOS	PubChem CID	S_{TAN}
Composto 1	11701241	1
Composto 2	57654171	0,81
Composto 3	11671139	0,56
Composto 4	14574745	0,83
Composto 5	5375528	1
Composto 6	13016529	0,83
Composto 7	15167264	1
Composto 8	12075518	0,81
Composto 9	153704416	1
Composto 10	153704415	1
Composto 11	102372082	1
Composto 12	153704414	1
Composto 13	4367260	1
Composto 14	5463312	0,98
Composto 15	87253260	0,81
Composto 16	11287766	0,98
Composto 17	20513050	0,98
Composto 18	87241372	0,79
Composto 19	102372083	0,64
Composto 20	53888304	0,96

4. CONCLUSÕES

No atual cenário mundial que envolvem drogas ilícitas e, especificamente, os Canabinoides Sintéticos, a busca rápida e otimizada por novas substâncias com atividade colorimétrica para indicação presuntiva dessas drogas, torna-se extremamente relevante no combate ao tráfico e na garantia da lei e da ordem. Apesar do espaço químico ser considerado infinito, o espaço colorimétrico dos compostos é limitado, e sua exploração constitui em uma tarefa muito difícil, utilizando apenas métodos convencionais. Posto isso, a Triagem Virtual surge como

ferramenta inovadora na descoberta de novos compostos com atividade colorimétrica, avaliando *in silico* grandes bibliotecas estruturais contra um alvo químico específico, favorecendo assim a redução de esforços financeiros, infraestrutura e tempo envolvidos no processo de descoberta de novas químioestruturas.

A filtragem molecular preliminar em uma biblioteca virtual permitiu identificar 13 (treze) compostos com similaridade ao composto de referência 4-DMAB, e 20 (vinte) compostos com similaridade a substância de referência 2,4-DNPH. Ao passo que a aplicação de metodologias de comparação por similaridade, reduziu ainda mais o número de compostos, o que resultou em 09 (nove) compostos com similaridade ao 4-DMAB e 18 (dezoito) compostos com similaridade ao 2,4-DNPH.

Assim, as moléculas de maior pontuação podem ser validadas em testes experimentais, reduzindo a quantidade de experimentação desnecessária realizada em laboratório, concentrando-se apenas nos candidatos mais promissores e reduzindo a relação custo/benefício nos programas de desenvolvimento de novos agentes colorimétricos. Além disso, o processo de Triagem Virtual baseada em ligantes fornece enriquecimento significativo sobre a seleção de moléculas em bases de dados, além de explorar uma ampla variedade de estruturas moleculares e características químicas em um curto período de tempo, permitindo a identificação de compostos com atividade colorimétrica em áreas do espaço químico que não podem ser consideradas em outros métodos.

AGRADECIMENTOS

Agradecimento às agências de fomento FAPESPA (Termo de Outorga n° 046/2021), CAPES e CNPq (Processo 307370/2022), à Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA) e a Polícia Científica do Pará - Superintendência Regional II em Marabá. Essa colaboração foi fundamental para o desenvolvimento e sucesso deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] J.A. Velho; G.C. Geiser; A. Espindula. Ciências forenses: *uma introdução às principais áreas da criminalística moderna*, Millennium, Brasil (2017) 505.
- [2] R.C.A. Isaacs. A structure-reactivity relationship driven approach to the identification of a color test protocol for the presumptive indication of synthetic cannabimimetic drugs of abuse. *Forensic Science International* **242**: 135-141 (2014).
- [3] A.T. Bruni; J.A. Velho; M.F. Oliveira. *Fundamentos de química forense*, Millenium, Brasil (2019) 380.
- [4] M.O.B. Arrieiro. Desenvolvimento de um método de triagem simples e rápido para determinação da Eutilona em amostras forenses. *Dissertação de Mestrado*, Departamento de Química, Universidade Federal dos Vales do Jequitinhonha e Mucuri (2021).
- [5] A.M.N. Santos. Análise Toxicológica de canabinóides sintéticos em contexto forense. *Dissertação de Mestrado*, Departamento de Ciências Farmacêuticas, Universidade Fernando Pessoa (2018).
- [6] Polícia Científica do Pará, Laboratório da Superintendência Regional II em Marabá, (2024). <https://www.pc.pa.gov.br/delegacias#>. Acesso em 03/05/2023).
- [7] S.J. Macalino; V. Gosu; S. Hong; S. Choi. Role of computeraided drug design in modern drug discovery. *Archives Pharmacol Research* **38**: 1686-1701 (2015).
- [8] Y. Chen; J. Kirchmair. Cheminformatics in natural product-based drug discovery. *Molecular Informatics* **39**: 2000-2171 (2020).
- [9] V. Auwarter; S. Dresen; W. Weinmann; M. Muller; M. Putz; N. Ferreirós. 'Spice' and other herbal blends: harmless incense or cannabinoid designer drugs? *Journal of Mass Spectrometry* **44**: 832-837 (2009).
- [10] H. Durmus; S. Durmazel; A. Üzer; B. Gökdere; E. Erçag; R. Apak. Colorimetric Determination of (Aminoalkyl)índole-containing Synthetic Cannabimimetics. *Analytical Sciences* **10**: 1419-1425 (2018).
- [11] <https://pubchem.ncbi.nlm.nih.gov/>. Acesso em 02/10/2023.
- [12] Open Moleculares. Osires Data Warrior. Disponível em: <http://www.openmoleculares.org/datawarrior>. Acesso em 09/06/2023.
- [13] A. A. Shelat; R.K. Guy. Scaffold composition and biological relevance of screening libraries. *Nature Chemicals Biology* **3**: 442-446 (2007).
- [14] A. Cereto-Massagué; M. J. Ojeda; C. Valls; M. Mulero; S. Garcia-Vallvé; G.Pujadas. Molecular fingerprint similarity search in virtual screening. *Methods* **71**: 58-63 (2015).
- [15] G. Glenner; R.D. Lillie. The histochemical demonstration of índole derivatives by the post-coupled p-dimethylaminobenzylidene reaction. *Journal of Histochem Cytochem* **5**: 279-296 (1957).
- [16] N.R. Campbell. The use of 2,4-dinitrophenylhydrazine as a reagent for carbonyl compounds. *The Analyst* **61**: 723-756 (1936).
- [17] O.L. Brady; G.V. Elsmie. O uso de 2:4-dinitrofenilidrazina como reagente para aldeídos e cetonas. *Analista* **51**: 77-78 (1926).
- [18] J. Tristão; M.M. dos Santos; J. V. Paveski; M. Bolzan; T. Foppa. Validação de metodologia analítica para realização de teste preliminar de identificação de maconha (*Cannabis sativa* spp) para fins forenses. *Extensão em Foco* **6**: 2317-9791 (2018).
- [19] R.P. Rodrigues; S.P. Mantoani; J.R. de Almeida; F.R. Pinsetta; E.P. Semighini; V.B. da Silva; C.H.P. da

- Silva. Estratégias de triagem virtual no Planejamento de fármacos. *Revista Virtual de Química* **4**: 739-776 (2012).
- [20] S. Kim; J. Chen; T. Cheng; A. Gindulyte; J. He; S. He; Q. Li; B.A. Shoemaker; P.A. Thiessen. PubChem update: improved access to chemical data. *Nucleic acids research* **47**: 1102-1109 (2019).
- [21] Q. Li; T. Cheng; Y. Wang; S.H. Bryant. PubChem as a public resource for drug discovery. *Drug Discovery Today* **15**: 23-24 (2010).
- [22] T.I. Oprea; J. Gottfries. Chemography: the art of navigating in chemical space. *Journal of Combinatorial Chemistry* **3**: 157-166 (2001).
- [23] T.K. Nguyen; C.B. Lorenz; V.D. Ruud; J.L. Reymond. Classification of organic molecules by molecular quantum numbers. *ChemMedChem* **4**: 1805-1805 (2009).
- [24] P.D. Leeson; S.A. ST-Gallay; M.C. Wenlock. Impact of ion class and time on oral drug molecular properties. *Medicinal Chemistry Communications* **2**: 91-105 (2011).
- [25] T. Sander; J. Freyss; M. von Korff; C. Rufener. Data Warrior: an open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling* **23**: 460-473 (2015).
- [26] J.B. Russel. *Química geral*. Makron. Brasil (1996) 516-555.
- [27] J. Arnott; S. Planey. The influence of lipophilicity in drug Discovery and design. *Expert Opinion on Drug Discovery* **7**: 863-875 (2012).
- [28] Z.Y. Yang; J.H. He; A.P. Lu; T.J. Hou; D.S. Cao. Application of Negative Design a More Desirable Virtual Screening Library. *Journal of Medicinal Chemistry* **63**: 4411-4429 (2020).
- [29] J. Kelder; P.D. Grootenhuis; D.M. Bayada; L.P. Delbressine; J.P. Ploemen. Polar Molecular Surface as a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs. *Pharmaceutical Research* **16**: 1514-1519 (1999).
- [30] D.F. Veber; S.R. Johnson; H.Y. Cheng; B.R. Smith; K.W. Ward; K.D. Kopple. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry* **45**: 2615-2623 (2002).
- [31] H. Yang; L. Sun; W. Li; G. Liu; Y. Tang. In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Frontiers in Chemistry* **6**: 129 (2018).
- [32] J. Velkoborsky; D. Hoksza. Scaffold analysis of PubChem database as background for hierarchical scaffold-based visualization. *Journal of Cheminformatics* **8**: 74, (2016).
- [33] A. Schuffenhauer; P. Ertl; S. Roggo; S. Wetzel; M.A. Koch; H. Waldmann. The scaffold tree-visualization of the scaffold universe by hierarchical scaffold classification. *Journal of Chemical Information and Modeling* **1**: 47-58 (2007).
- [34] D. Bajusz; A. Rácz; K. Héberger. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?. *Journal of Cheminformatics* **7**: 20 (2015).
- [35] L. Kaufman; P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Interscience. England (1990) 1618-1620.
- [36] M.G. Maggiora; J. Bajorath. Chemical space networks: a powerful new paradigm for the description of chemical space. *Journal of Computer-Aided Molecular Design* **28**: 795-802 (2014).
- [37] R. Kunimoto; J. Bajorath. Combining similarity searching and network analysis for the identification of active compounds. *ACS Omega* **3**: 3768-3777 (2018).
- [38] B.F. Domingues. 3D-Pharma: uma ferramenta para triagem virtual baseada em fingerprints de farmacofóros. *Tese de Doutorado*, Departamento de bioquímica e imunologia, Universidade Federal do Rio Grande do Sul, (2001).
- [39] A. Rácz; D. Bajusz; K. Héberger. (2018). Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. *Journal of Cheminform* **10**: 48 (2018).
- [40] I.T. Jolliffe. Principal Component Analysis, Series: Springer Series in Statistics. *Springer* **2**: 28 (2002).
- [41] L.O. Magalhães. Desenvolvimento de métodos quimiométricos para triagem de novas substâncias psicoativas em selos utilizando técnicas espectroscópicas na região do infravermelho. *Tese de Doutorado*, Departamento de Química, Universidade de Brasília, (2019).
- [42] D. Granato; J.S. Santos; G.B. Escher; B.L. Ferreira; R. Maggio. Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Journal Trends in Food Science and Technology* **72**: 83-90 (2018).
- [43] S. Ekins; J. Mestres; B. Testa. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology* **152**: 9-20 (2007).
- [44] B.O. Gohlke; T. Overkamp; A. Richter; P.T. Daniel; B. Gillissen; R. Preissner. 2D and 3D similarity landscape analysis identifies PARP as a novel off-target for the drug Vatalanib. *BMC Bioinformatics* **16**: 308 (2015).
- [45] D. Giganti; H. Guillemain; J.L. Spadoni; M. Nilges; J.F. Zagury; M. Montes. Comparative evaluation of 3D virtual ligand screening methods: impact of the molecular alignment on enrichment. *Journal of Chemical Information and Modeling* **50**: 992-1004 (2010).
- [46] K. Santana; L.D. do Nascimento; A.L. e Lima; V. Damasceno; C. Nahum; R.C. Braga; J. Lameira. Applications of virtual screening in bioprospecting: facts, shifts, and perspectives to explore the chemo-structural

diversity of natural products. *Frontiers in Chemistry* **9**: 662-688 (2021).

[47] S. Riniker; G.A. Landrum. Similarity maps—a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminform* **5**: 43 (2013).

[48] Y. Hu; D. Stumpfe; J. Bajorath. Exploração computacional de estruturas moleculares em química medicinal. *Journal of Medical Chemistry* **5**: 17-46 (2016).

[49] B.F.J. Manly. Multivariate statistical methods: a primer. *Chapman & Hall* **2**: 215 (1994).