



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS FARMACÊUTICAS

DISSERTAÇÃO DE MESTRADO

Rafael Campos Vieira

MÉTODOS DE DATA SCIENCE E DATA ANALYTICS PARA O DESENVOLVIMENTO DE NOVOS FÁRMACOS

Linha de Pesquisa: Desenvolvimento e qualidade de produtos e serviços farmacêuticos

Pós-Graduação em Ciências Farmacêuticas

Orientador: Dr. João Batista Lopes Martins

Co-orientadora: Dra. Érica Cristina Moreno Nascimento

Brasília, 2025

RAFAEL CAMPOS VIEIRA

**MÉTODOS DE DATA SCIENCE E DATA ANALYTICS
PARA O DESENVOLVIMENTO DE NOVOS FÁRMACOS**

Dissertação de Mestrado apresentada ao Programa de
Pós-Graduação em Ciências Farmacêuticas
da Universidade de Brasília como requisito para obtenção
do título de Mestre em Ciências Farmacêuticas.

Orientador: Dr. João Batista Lopes Martins

Co-orientadora: Dra. Érica Cristina Nascimento Moreno

Linha de Pesquisa: Desenvolvimento e qualidade de produtos e serviços farmacêuticos

Brasília, 2025

Resumo

Este trabalho investiga a aplicação de algoritmos de machine learning no campo do drug- design, com ênfase na descoberta de novas moléculas candidatas a inibidores da enzima acetilcolinesterase humana (hAChE), um alvo terapêutico relevante no tratamento da doença de Alzheimer. Esse estudo visa o uso de canabinóides e canabinóides raros como moléculas de partida, com o objetivo não apenas de identificar compostos com potencial terapêutico, mas também de aprimorar o processo de virtual screening por meio de técnicas avançadas de data science. Para isso, foi desenvolvido um banco de dados contendo propriedades físico-químicas dos compostos, que inclui descritores ADMET (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade) e estrutura eletrônica, utilizando os softwares SwisADME e Gaussian, respectivamente.

Foram aplicados diferentes algoritmos modernos de machine learning, como k-means para clusteração, máquinas de vetores de suporte (do inglês *support vector machines*, SVM) para classificação, análise dos componentes principais (do inglês *principal component analysis*, PCA) para redução de dimensionalidade. Foram estruturadas e aplicadas modelos de redes neurais para a geração de novas moléculas, usando o conteúdo total de 310253 moléculas, coletadas em três bancos de dados de moléculas disponíveis. Para buscar e identificar padrões significativos relacionados com as propriedades físico-químicas, estruturais, eletrônicas e de ADMET das moléculas estudadas, foram empregadas técnicas de visão computacional para analisar as imagens das interações não covalentes (do inglês *non-covalent interactions*, NCI) e os mapas de potencial eletrostático (do inglês *molecular electrostatic potential*, MEP).

As interações dos complexos formados entre os resíduos do sítio ativo da hAChE e os compostos estudados, foram previstas por meio de simulações de interações receptor-ligante, aplicando a técnica de docking molecular, usando o software AutoDock. A identificação de parâmetros mais relevantes para atividade biológica dos candidatos foi obtida por a análise conjunta dos dados de ADMET e estrutura eletrônica, utilizando os algoritmos k-means e PCA.

Os resultados obtidos sugerem que a combinação de machine learning com docking molecular, análise de propriedades físico-químicas e de estrutura eletrônica, revelou ser uma ferramenta poderosa no desenvolvimento de novos candidatos a fármacos para o tratamento da doença de Alzheimer. O estudo também contribui para a evolução do processo de virtual screening, ao fornecer modelos robustos para a análise de dados relevantes para a atividade biológica e a seleção de moléculas candidatas com maior potencial terapêutico. Das moléculas estudadas, sete se destacaram, com o perrotetino apresentando perfil mais semelhante ao ligante de referência donepezil, mimetizando interações cátion-pi e exibindo score semelhante de -10,32 kcal/mol.

Palavras chave: aprendizado de máquina, redes neurais artificiais, estrutura eletrônica, canabinóides raros, docking molecular.

Abstract

This work investigates the application of machine learning algorithms in the field of drug design, with an emphasis on the discovery of new candidate molecules as inhibitors of the human acetylcholinesterase enzyme (hAChE), a relevant therapeutic target in the treatment of Alzheimer’s disease. This study aims to use cannabinoids and minor cannabinoids as starting molecules, with the goal not only of identifying compounds with therapeutic potential, but also of improving the virtual screening process through advanced data science techniques. To this end, a database was developed containing physicochemical properties of the compounds, which includes ADMET descriptors (Absorption, Distribution, Metabolism, Excretion, and Toxicity) and electronic structure, using the software SwisADME and Gaussian, respectively.

Different modern machine learning algorithms were applied, such as k-means for clustering, support vector machines (SVM) for classification, and principal component analysis (PCA) for dimensionality reduction. Neural network models were structured and applied for the generation of new molecules, using the total content of 310253 molecules, collected from three available molecular databases. To search for and identify significant patterns related to the physicochemical, structural, electronic, and ADMET properties of the studied molecules, computer vision techniques were employed to analyze the images of non-covalent interactions (NCI) and the molecular electrostatic potential (MEP) maps.

The interactions of the complexes formed between the active site residues of hAChE and the studied compounds were predicted by means of receptor-ligand interaction simulations, applying the molecular docking technique, using the software AutoDock. The identification of the most relevant parameters for the biological activity of the candidates was obtained through the joint analysis of ADMET and electronic structure data, using the k-means and PCA algorithms.

The results obtained suggest that the combination of machine learning with molecular docking, physicochemical property analysis, and electronic structure evaluation has proven to be a powerful tool in the development of new drug candidates for the treatment of Alzheimer’s disease. The study also contributes to the advancement of the virtual screening process by providing robust models for the analysis of data relevant to biological activity and the selection of candidate molecules with greater therapeutic potential. Among the studied molecules, seven stood out, with perrotetineno presenting a profile most similar to the reference ligand donepezil, mimicking cation- π interactions and exhibiting a similar score of -10.32 kcal/mol.

Keywords: machine learning, artificial neural networks, electronic structure, minor cannabinoids, molecular docking.

Dedicatória

Dedico este trabalho, primeiramente, aos meus pais, Joilson Campos e Regina do Couto Campos de Jesus, que tanto investiram em mim desde os meus primeiros anos de vida. Recordo-me com alegria e emoção dos inúmeros livros que pedia à minha amada mãe, e ela, sem hesitar, me concedia. E, quando não podia, fazia o possível para me presentear em uma oportunidade mais favorável.

O mesmo vale para o meu querido pai, que desde pequeno me incentivou a desenvolver o pensamento crítico e científico, sendo o grande responsável por despertar em mim a fascinação pela farmacologia. Lembro-me, ainda criança, folheando o livro de farmacologia do Dale que ele possuía, ou a segunda edição do de bioquímica do Lehninger — sem compreender muito, mas já profundamente encantado pela ciência.

Vocês sempre me proporcionaram as melhores oportunidades de estudo, me deram todo o suporte necessário e foram fundamentais para que eu chegasse até aqui.

À minha amada irmã, dedico também estas linhas, pois sempre foi para mim um exemplo de estudo, dedicação, disciplina, amor à família e zelo.

Dedico este trabalho, ainda, à minha esposa, Thainá Campos Vieira. Como não me emocionar ao lembrar que, se não fosse você, eu talvez não tivesse tido coragem de enfrentar este desafio? Você, que tantas vezes acreditou mais em mim do que eu mesmo. Saiba que esta conquista não é só minha — é nossa.

Estendo esta dedicatória também aos demais familiares: meus avós, tios, tias e primos — tanto os que ainda estão conosco quanto os que já partiram —, bem como àqueles tios, tias, primos e à mãe que a vida me presenteou e que hoje tenho o privilégio de chamar de família. A todos vocês, o meu muito obrigado.

Por fim, e acima de tudo, agradeço ao meu Senhor Jesus Cristo. A Ele, toda a honra e toda a glória.

“Senhor, dai-me inteligência para conhecer-vos, diligência para procurar-vos e sabedoria para vos encontrar.” (Santo Tomás de Aquino).

Agradecimentos

Agradeço, primeiramente, a Deus pelo dom da vida e por me conceder os meios necessários para chegar até aqui. À minha família, por todo o suporte, apoio e amor incondicional ao longo desta caminhada.

À CAPES e à FADPDF, pelo auxílio financeiro fundamental ao desenvolvimento deste trabalho. À Universidade de Brasília e ao Programa de Pós-Graduação em Ciências Farmacêuticas, por todo o apoio, tanto institucional quanto financeiro, que me proporcionaram a oportunidade de participar de congressos nacionais e internacionais ao longo desta jornada.

Aos meus orientadores, professor João e professora Érica, por toda a paciência, orientação e confiança depositadas em meu trabalho.

Ao corpo institucional e técnico de servidores da Faculdade de Saúde e do Instituto de Química, em especial às funcionárias da limpeza do IQ, por sempre manterem o laboratório impecável, garantindo as condições ideais para o desenvolvimento das nossas atividades.

A toda a família LQC, em especial aos amigos do LQC-Bio, por todos os ensinamentos, pela parceria, pelos momentos de descontração e também pelo apoio mútuo nos períodos de aflição e incertezas.

Aos meus irmãos de comunidade, por todas as orações, palavras de incentivo e motivações que me sustentaram nos momentos mais difíceis.

Lista de Figuras

1.1	Estágios da DA	13
1.2	Cadeia α da hAChE	16
1.3	Representação em 2D das estruturas dos AChEIs clássicos	18
1.4	Moléculas de partida para novos AChEIs	19
1.5	Representação em 2D da estrutura geométrica de alguns AChEI/BChEI Multi-targets	20
1.6	Moléculas extraídas de algas como compostos de partida	20
1.7	Principais canabinóides encontrados na <i>Cannabis</i>	21
1.8	Principais radulaninas	22
1.9	Esquema estrutural dos principais perrotetinenos	23
1.10	Canabinóides bibenzílicos	24
1.11	Estrutura química dos machaeriols	24
1.12	Estrutura química dos machaeridiols	25
1.13	Representação 2D das desmodianonas usadas no presente trabalho.	25
1.14	Principais araquidinas	26
3.1	Algoritmo GD bidimensional	32
3.2	Simulação algoritmo SVM	33
3.3	Máquina de vetor de suporte com kernel	37
3.4	Simulação da reconstrução dos dados via equação de projeção	42
3.5	Relação entre função de custo, componentes principais e autovalores	43
3.6	Representação rede neural artificial	45
3.7	Representação autoencodificador	46
3.8	Representação de uma operação de convolução	47
3.9	Representação operações em redes convolucionais	47
4.1	Protocolo de redocking	49
4.2	Exemplo HeatLQC	51
4.3	Exemplo PCAAnalytics	52
4.4	Esquema ilustrativa rede siamesa-NCIVISION	52
4.5	Erro durante treinamento do modelo NCIVISION	53
4.6	Representação esquemática da codificação do tipo <i>one-hot</i>	54
5.1	Interações DNP/GORGE	56
5.2	Estrutura eletrônica DNP	57
5.3	Interações PRT/GORGE	58
5.4	Estrutura eletrônica PRT	58
5.5	Interações DMDD/GORGE	59
5.6	Estrutura eletrônica DMDD	59
5.7	Interações MCRA/GORGE	60
5.8	Estrutura eletrônica MCRA	60
5.9	Interações DMDE/GORGE	61
5.10	Estrutura eletrônica DMDE	61
5.11	Interações MCDC/GORGE	62
5.12	Estrutura eletrônica MCDC	62
5.13	Interações OMCNB/GORGE	63
5.14	Estrutura eletrônica OMCNB	63
5.15	Interações D9CT/GORGE	64

5.16	Estrutura eletrônica D9CT	64
5.17	Comparação entre valores de variância explicada cumulativa	65
5.18	Análise espaços de características gerados pelo K-means	65
5.19	Análise com matrizes de correlação- <i>Heatmaps</i>	66
5.20	Análise similaridade de Tanimoto e distância invertida	67
5.21	Grad-CAM	68
5.22	Diagramas RDG de entrada	69
5.23	Mosaico dos <i>embeddings</i> com 64 dimensões gerados pelo modelo a partir dos diagramas RDG para os 32 compostos analisados no estudo.	70
5.24	Mapas de atenção gerados pelo modelo NCIVISION	71
5.25	Relação entre temperatura, τ e variância para o modelo generativo	72
5.26	Relação entre temperatura, τ , validade e unicidade para o modelo generativo	73
5.27	Moléculas geradas pelo modelo	74
7.3	Visualizações 3D dos resultados de estrutura eletrônica	104

Tabela com siglas das moléculas

Sigla	Nome
DNP	donepezil
LVD	levodopa
CMG	carmagerol
CNBRPL	canabiripsol
GLDC	glepidotina-C
OMCNB	O-metilcanabinol
PRT	perrotetinenol
THC	Δ^9 -tetraidrocanabinol
THCP	tetraidrocanabiforol
THCVM2	tetraidrocanabivarina-M2 (metabólito)
CNBA	ácido canabinólico
HHC	hidroxihelicanabigerol
CNBD	canabinodiol
ARAC1	araquidina-1
ARAC2	araquidina-2
MCDC	machaeridiol-C
MCDD	machaeridiol-D
RDLA	radulanina-A
RDLK	radulanina-K
RDLI	radulanina-I
RDLH	radulanina-H
RDLJ	radulanina-J
RDL59	radulanina-59 (metabólito)
DCBA	descarboxiamorfrutina-A
DMDD	desmodianona-D
DMDE	desmodianona-E
D9CT	Δ^9 - <i>cis</i> -tetraidrocanabinol
TCNT	<i>trans</i> -canabitríol
HD11	11-hidroxi- Δ^9 -hexahidrocanabinol
DDAA	dimetildescarboxiamorfrutina-A
MCRA	machaeriol-A
R6E	<i>rac</i> -6-epoxicanabigerol

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

CC198m Campos Vieira, Rafael
MÉTODOS DE DATA SCIENCE E DATA ANALYTICS PARA O
DESENVOLVIMENTO DE NOVOS FÁRMACOS / Rafael Campos Vieira;
orientador João Batista Lopes Martins; co-orientador Érica
Cristina Nascimento Moreno. Brasília, 2025.
p.

Dissertação(Mestrado em Ciências Farmacêuticas)
Universidade de Brasília, 2025.

1. machine learning. 2. artificial neural networks. 3.
data science. 4. minor-cannabinoids. 5. molecular docking.
I. Batista Lopes Martins, João, orient. II. Cristina
Nascimento Moreno, Érica , co-orient. III. Título.

Sumário

Lista de Figuras	5
Sumário	11
1 Introdução	13
1.1 Doença de Alzheimer	13
1.1.1 Patologia da proteína tau: Tauopatias	14
1.1.2 Neuroinflamação	14
1.1.3 Extresse oxidativo	14
1.1.4 Acúmulo de metais	14
1.1.5 Outros fatores com potencial contribuição na DA	15
1.1.6 Hipótese colinérgica	15
1.1.7 Hipótese da cascata amiloide (A- β)	16
1.2 Terapias farmacológicas para as doenças DA e DP	17
1.2.1 AChEI/BChEI Single-target	17
1.2.2 Compostos de partida: moléculas naturais e tacrina	18
1.2.3 AChEI/BChEI Multi-target	19
1.2.4 Metabolitos de algas	20
1.2.5 Antagonistas do receptor NMDA	20
1.3 Canabinóides e canabinóides raros como potenciais agentes terapêuticos em doenças neu- rodegenerativas	21
1.4 Gênero <i>Radula</i>	22
1.4.1 Radulaninas	22
1.4.2 Canabinóides bibenzílicos	23
1.5 Gênero <i>Machaerium</i>	24
1.6 Gênero <i>Desmodium</i>	25
1.7 Gênero <i>Arachis</i>	26
2 Objetivos	27
2.1 Objetivo	27
2.2 Objetivos Específicos	27
3 Referencial Teórico	28
3.1 Teoria do Funcional de Densidade- DFT	28
3.2 Índices de Reatividade Química derivados da teoria do funcional de densidade	28
3.3 Interações não-covalentes	29
3.4 Gradiente de densidade reduzida-RDG	30
3.5 Docking Molecular	30
3.6 Métodos em Data Science	31
3.6.1 Gradiente descendente como método central de otimização em machine learning	31
3.7 Algoritmos de classificação	32
3.7.1 Máquinas de vetores de suporte	32
3.7.2 Análise discriminante gaussiana-GDA	37
3.8 Algoritmos de clusterização	38
3.8.1 Kmeans	38
3.9 Visão Computacional	39
3.9.1 Introdução	39

3.9.2	Métodos utilizados: GLCM	39
3.9.3	Métodos utilizados: Filtro de Sobel	40
3.10	Métodos para tratamento dos dados	40
3.10.1	Análise dos componentes principais-PCA	40
3.11	Redes neurais artificiais	44
3.11.1	Introdução	44
3.11.2	Retropagação	45
3.11.3	Autoencodificador variacional- VAE	46
3.11.4	Redes neurais convolucionais	46
4	Metodologia	48
4.1	Protocolo de docking	48
4.2	Cálculos de estrutura eletrônica	49
4.3	Coleta de dados ADMET e de estrutura eletrônica	49
4.4	Algoritmos desenvolvidos e implementados	49
4.4.1	Ferramentas usadas	49
4.4.2	Software HeatLQC	50
4.4.3	PCA Analytics	51
4.4.4	Rede neural siamesa para similaridade via Imagens MEP	52
4.4.5	Rede neural generativa para geração de moléculas-VAELSTM	54
5	Resultados e Discussão	56
5.1	Docking e estrutura eletrônica	56
5.2	Data analytics	65
5.2.1	Principais PCs para o conjunto de dados estudado	65
5.3	Redes neurais artificiais	67
5.3.1	Rede neural siamesa-NCIVISION	67
5.3.2	Autoencodificador variacional generativo acoplado a células de memória de curta e longa duração-VAELSTM	72
6	Conclusões	75
7	APÊNDICES	98
7.1	Tabelas suplementares	98
7.2	Visualizações dos resultados dos cálculos de estrutura eletrônica	103
7.3	Patente de software	105

Capítulo 1

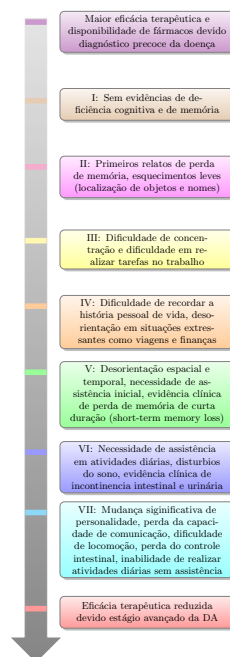
Introdução

1.1 Doença de Alzheimer

A doença de Alzheimer (DA) é caracterizada pelo declínio cognitivo progressivo, afetando majoritariamente a população idosa com 60 anos ou mais (Tzioras *et al.*, 2023). O progresso da doença é normalmente dividido em sete fases (Mirakhori *et al.*, 2022; Hemalatha; Renukadevi, 2021; Mehmood *et al.*, 2020; Galasko *et al.*, 1991): I-Sem declínio cognitivo clínico; II-Início do declínio cognitivo; III-Leve intensificação do declínio cognitivo; IV-Declínio cognitivo moderado; V-Declínio cognitivo moderadamente severo; VI-Declínio cognitivo severo; VII-Declínio cognitivo consideravelmente severo (demência).

A demência é o estágio final da DA, sendo a palavra demência de origem latina, *demens*, que significa “sem mente” (Ahmed *et al.*, 2021). A DA afeta amplas regiões do córtex cerebral e do hipocampo, com as primeiras alterações geralmente ocorrendo no tecido cerebral que reveste o lobo frontal e temporal (Masters *et al.*, 2015). A Organização Mundial da Saúde estima que 1 bilhão de pessoas tinham 60 anos ou mais em 2020, e que esta faixa etária irá dobrar para 2.1 bilhões de pessoas em 2050, com dois terços desta população vivendo em países de baixa ou média renda percapita (WHO, 2022).

Figura 1.1: Estágios da DA. Adaptado do site Alzheimer diseases (Alzheimer’s Disease, 2019)



Compreender melhor a doença é fundamental para o desenvolvimento de novas abordagens terapêuticas voltadas ao tratamento dos sintomas da DA, bem como para a busca de sua cura. A etiologia da DA ainda não é completamente compreendida, em razão de seu caráter multifatorial. As estratégias farmacoterapêuticas atualmente disponíveis atuam majoritariamente sobre os sintomas cognitivos, não havendo,

até o momento, fármacos seguros e eficazes capazes de prevenir ou reverter a progressão da doença (Stanciu *et al.*, 2019). A seguir, serão abordadas as principais teorias relacionadas à origem e à evolução da doença, as quais fundamentam os estudos voltados ao desenvolvimento de fármacos para o tratamento da DA.

1.1.1 Patologia da proteína tau: Tauopatias

As tauopatias são doenças neurodegenerativas caracterizadas pela formação de emaranhados neurofibrilares intracelulares (*neurofibrillary tangles* – NFTs), compostos pela proteína tau hiperfosforilada. A presença de NFTs no cérebro, em conjunto com a agregação da proteína beta-amiloide ($A\beta$), constituem um dos principais marcadores patológicos da DA. Tauopatias apresentam elevada associação com quadros de demência, especialmente nos estágios mais avançados da doença (Chong *et al.*, 2018).

Em condições fisiológicas, a proteína tau exerce papel fundamental na manutenção da integridade dos neurônios do sistema nervoso central (SNC), atuando na estabilização dos microtúbulos e na regulação das sinalizações sinápticas. Contudo, quando sofre hiperfosforilação, a proteína tau perde sua capacidade de se ligar aos microtúbulos, promovendo a instabilidade dessas estruturas e contribuindo para a degeneração neuronal. Dessa forma, a hiperfosforilação da proteína tau é considerada um fator-chave na fisiopatologia das tauopatias (Alavi Naini; Soussi-Yanicostas, 2015; Takashima, 2008; Iqbal *et al.*, 2013).

Outro mecanismo patológico relevante nas tauopatias é a oligomerização da proteína tau, que pode induzir sua agregação antes mesmo da formação dos NFTs (Clavaguera *et al.*, 2009; Fá *et al.*, 2016). Estudos *in vivo* e *in vitro* demonstram que a presença desses oligômeros nas sinapses é altamente tóxica para os neurônios, provocando disfunções sinápticas precoces e contribuindo para a neurodegeneração característica da doença de Alzheimer, ainda na ausência de emaranhados neurofibrilares completamente formados (Tai *et al.*, 2014).

1.1.2 Neuroinflamação

A neuroinflamação envolve a ativação de células da glia, como a microglia e os astrócitos. As microglias são células imunes residentes no SNC, já os astrócitos, que fornecem suporte metabólico e funcional aos neurônios. A resposta imune desencadeada pela agregação de proteínas patológicas, como beta-amiloide ($A\beta$) e tau, parece desempenhar um papel importante na progressão da doença de Alzheimer (Meda *et al.*, 2001; Medeiros; LaFerla, 2013). No entanto, a distribuição temporal e espacial da resposta inflamatória, assim como sua relevância direta para a progressão da doença, ainda precisa ser analisada de forma mais sistemática e abrangente (Trejo-Lopez *et al.*, 2022).

1.1.3 Estresse oxidativo

O estresse oxidativo é um processo bioquímico que contribui tanto para o envelhecimento cognitivo quanto para o desenvolvimento de doenças neurodegenerativas. No contexto neuronal, o estresse oxidativo está associado à disfunção mitocondrial, ao dano ao DNA e a alterações epigenéticas relacionadas ao envelhecimento cerebral e ao surgimento da DA (Ionescu-Tucker; Cotman, 2021).

Um estudo clássico conduzido por Hensley e colaboradores, em 1994, conseguiu modelar sinteticamente a agregação de proteínas $A\beta$, com base na geração de radicais livres induzida pelo próprio peptídeo. Esse experimento demonstrou a possível contribuição de espécies reativas no processo de agregação e neurotoxicidade das proteínas $A\beta$ (Hensley *et al.*, 1994).

Outra evidência importante é o acúmulo de íons de ferro no cérebro de indivíduos com DA. Esses íons de ferro podem participar de reações de oxidorredução (redox), resultando na geração de espécies radiculares, o que intensifica os danos celulares e contribui para a progressão das doenças neurodegenerativas (Smith *et al.*, 1997).

1.1.4 Acúmulo de metais

O acúmulo específico dos metais ferro e alumínio foi observado em emaranhados neurofibrilares em cérebros de indivíduos com DA (Good *et al.*, 1992). Como mencionado anteriormente, os íons ferro acumulados no cérebro contribuem significativamente para o estresse oxidativo neuronal, enquanto o alumínio, em concentrações elevadas, pode induzir à neurotoxicidade (Muma; Singer, 1996).

Estudos epidemiológicos identificaram uma associação geográfica entre a incidência de DA e a presença de alumínio na água potável (Martyn *et al.*, 1989; Neri; Hewitt, 1991). No entanto, embora essas correlações sejam sugestivas, é improvável que estabeleçam uma relação causal plausível, dada a natureza

multifatorial da DA. Ainda assim, a presença desses metais em concentrações elevadas no tecido cerebral pode contribuir para a progressão da doença.

1.1.5 Outros fatores com potencial contribuição na DA

Além das principais hipóteses abordadas nas seções anteriores, outros fatores e patologias têm demonstrado relevância clínica para o diagnóstico e a progressão da DA, incluindo:

Anormalidades vasculares: Comprometem o fornecimento de nutrientes e a remoção de produtos metabólicos, levando à ocorrência de microinfartos e à ativação das células gliais (Daulatzai, 2017).

Barreira hematoencefálica: A integridade e funcionalidade dessa barreira são essenciais para proteger o SNC; sua disfunção pode permitir a entrada de agentes neurotóxicos (Huang *et al.*, 2020).

Dislipidemias: Alterações no metabolismo do colesterol têm sido associadas a quadros clínicos de DA (Reitz, 2013).

Diabetes mellitus tipo 2: A resistência à insulina no cérebro afeta a função neuronal e a plasticidade sináptica (Arnold *et al.*, 2018).

Microbiota intestinal: Alterações na composição da microbiota podem modular a inflamação sistêmica e influenciar a neuroinflamação (Hilgert; Dias, 2023).

Erro na codificação dos Alelos APOE4: Associado ao aumento da produção e à redução da depuração da proteína A β (Blumenfeld *et al.*, 2024).

Comorbidades neuropatológicas: Incluem doença de Corpus de Lewy, encefalopatia TDP-43 relacionada à idade (LATE), encefalopatia traumática crônica (ETC) e astrogliopatia tau (ARTAG) (Wu *et al.*, 2025).

1.1.6 Hipótese colinérgica

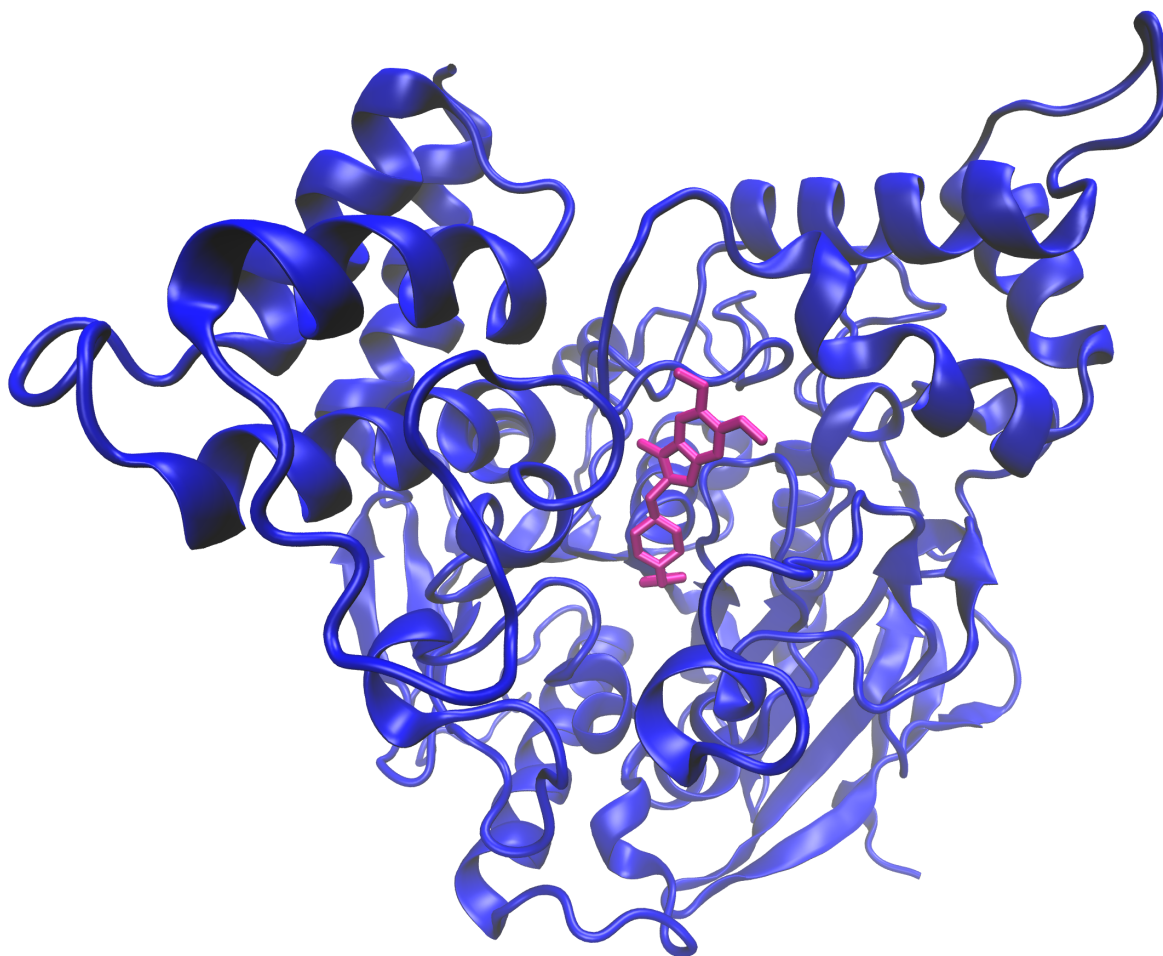
A hipótese colinérgica, a mais antiga teoria sobre a patogênese da DA, propõe que a falha na síntese do neurotransmissor acetilcolina (ACh) e sua hidrólise, resultando na alteração do gradiente de concentração desse neurotransmissor, constituem os principais fatores etiológicos da enfermidade. Essa diminuição está diretamente associada à atrofia dos neurônios colinérgicos e ao consequente declínio cognitivo acelerado, especialmente em doenças neurodegenerativas relacionadas ao envelhecimento (Chen *et al.*, 2022). A ACh é um neurotransmissor excitatório endógeno essencial, desempenhando papel fundamental nos processos de aprendizagem e memória (Huang *et al.*, 2022). Sua descoberta está intimamente ligada aos estudos pioneiros de Henry Dale e Otto Loewi, agraciados com o Prêmio Nobel de Fisiologia e Medicina em 1936.

Em 1914, as investigações de Henry Dale, resultaram no isolamento de uma substância com propriedades estimulantes a partir do fungo *Claviceps purpurea* (esporão-do-centeio), antecipando a caracterização da ACh como neurotransmissor (Dale, 1914).

Em 1921, Loewi conduziu um experimento seminal no qual isolou dois corações de sapos, retirando os nervos de um deles. Ambos foram imersos em solução de Ringer, e o nervo vago do coração intacto foi estimulado por alguns minutos. Em seguida, a solução foi transferida para o coração desnervado, provocando uma redução significativa na frequência cardíaca, similar ao efeito observado no coração com inervação preservada. A substância responsável por esse efeito foi inicialmente denominada *vagusstoff* — termo alemão que significa “substância do nervo vago” — e posteriormente identificada como acetilcolina (Loewi, 1921).

Com base na hipótese colinérgica, foram desenvolvidos os primeiros fármacos para o tratamento da DA, ainda amplamente utilizados na prática clínica. Esses agentes terapêuticos, como donepezil, rivastigmina e galantamina, atuam inibindo a atividade enzimática da acetilcolinesterase humana (hAChE) sobre a ACh, promovendo o aumento da disponibilidade sináptica do neurotransmissor. Além da DA, esses inibidores demonstram eficácia em outras doenças neurodegenerativas, como a doença de Parkinson (DP).

Figura 1.2: Cadeia α da hAChE (estrutura depositada no *Protein Data Bank* sob o código 6O4W) em azul escuro complexada com o inibidor donepezil (estrutura em tubo) em magenta.



1.1.7 Hipótese da cascata amiloide ($A\beta$)

A hipótese da cascata amiloide, proposta por John Hardy e David Allsop em 1991 (Hardy; Allsop, 1991), sugere que a desregulação metabólica da proteína precursora amiloide (do inglês *amyloid precursor protein*, APP), localizada no cromossomo 21, constitui o fator causal primário no desenvolvimento da DA. Essa hipótese fundamenta-se na identificação de mutações no gene APP, que levam ao acúmulo anormal do peptídeo beta-amiloide ($A\beta$), contribuindo significativamente para a patogênese da doença (Liu *et al.*, 2019). De acordo com essa teoria, o acúmulo progressivo do peptídeo $A\beta$ — derivado da clivagem anormal da APP — atua como o principal fator etiológico da DA (Masters *et al.*, 2015). A agregação dos peptídeos $A\beta$ resulta na formação de placas senis insolúveis, que se depositam nos espaços extracelulares do tecido cerebral e nas paredes vasculares.

Simultaneamente, a polimerização anormal da proteína tau — cuja hipótese será discutida a seguir — leva à formação de emaranhados neurofibrilares nos neurônios. A interação entre estas duas patologias, placas amiloides e emaranhados tau, contribui de forma sinérgica para a progressão e severidade dos danos neurodegenerativos observados na DA. (Masters *et al.*, 2015).

O caráter complexo e multifatorial da DA, indica a necessidade de abordagens integrativas para o diagnóstico, tratamento e prevenção da doença. Na próxima seção serão abordados os principais fármacos utilizados atualmente para o tratamento da doença, bem como as principais moléculas que tem sido utilizadas como compostos de partida para o desenvolvimento de novos inibidores da enzima AChE (AChEI).

1.2 Terapias farmacológicas para as doenças DA e DP

1.2.1 AChEI/BChEI Single-target

Primeiramente serão discutidos a tacrina (THA) e os três fármacos comumente usados para o tratamento da DA: donepezil (DNP), rivastigmina e galantamina (Figura 1.3). A tacrina, ou tetrahydroaminoacridina, foi o primeiro fármaco aprovado pela FDA para o tratamento da DA (Crismon, 1994). Sua síntese inicial ocorreu em 1945, por Adrien Albert, na Universidade de Sydney, Austrália, como parte de um projeto de busca por agentes antibacterianos para tratar feridas infectadas em soldados durante a Segunda Guerra Mundial (Jarrott, 2017). No entanto, a substância não demonstrou atividade em cepas de bactérias em testes *in vitro*.

A descoberta de seu potencial atividade inibitória na hAChE veio posteriormente, quando William Summers, em 1980, formulou a hipótese sobre o possível benefício da tacrina nos estágios iniciais da demência na DA (Summers *et al.*, 1981). Após estudos mais abrangentes que corroboraram os efeitos positivos da droga, Summers obteve a patente da tacrina para uso no tratamento da DA em 1989 (Summers, 1989). A tacrina atua como um inibidor reversível da hAChE, mas também apresenta atividade farmacológica em outras enzimas, como na butirilcolinesterase (BChE), bem como nos receptores de N-metil-D-aspartato (NMDAR) e receptores ionotrópicos.

Apesar de ter sido o primeiro fármaco aprovado para o tratamento da DA, seu uso foi limitado devido a seus efeitos colaterais, especialmente sua hepatotoxicidade (Watkins *et al.*, 1994). Devido a essa limitação, e com o desenvolvimento de AChEIs mais seguros, a tacrina foi gradualmente retirada do mercado em diversos países, incluindo os Estados Unidos em 2013 (Marucci *et al.*, 2021). Contudo, seu desenvolvimento abriu caminho para a criação de novos AChEIs, sendo utilizada até os dias atuais como composto de partida em estudos de drug design.

O DNP é um dos principais fármacos utilizados no tratamento sintomático da DA, com perfil de segurança superior ao da THA. Atua como um inibidor reversível e não competitivo da enzima hAChE, promovendo um aumento significativo na concentração extracelular de acetilcolina no córtex cerebral e no hipocampo (Wilkinson, 1999). Esse composto foi desenvolvido pela farmacêutica japonesa Eisai, sob a liderança do químico e farmacologista Hachiro Sugimoto. Após a conclusão bem-sucedida dos estudos clínicos de fase I no Japão, em 1989, o DNP foi aprovado pela *Food and Drug Administration*-FDA, dos Estados Unidos em novembro de 1996 (Takeda; Tagami, 2020).

O uso contínuo do DNP exerce um efeito modulador positivo sobre os receptores colinérgicos nicotínicos (nAChR), o que contribui para seu potencial mecanismo de neuroproteção (Takada-Takatori *et al.*, 2008; Kume; Takada-Takatori, 2018). Do ponto de vista estrutural, apresenta um núcleo de piperidina central, ligado a dois grupos funcionais: um grupamento benzila e um anel de indenona dissubstituído com dois grupos metóxi (Figura 1.3). Entre os principais efeitos adversos associados ao seu uso estão náuseas e vômitos, anorexia, fadiga, insônia e câibras musculares (Pratt *et al.*, 2002; Shintani; Uchida, 1997; Adlimoghaddam *et al.*, 2018; Bruera *et al.*, 2007).

A rivastigmina é um inibidor não competitivo e pseudo-irreversível das enzimas hAChE e BChE (Alcolea-Palafox *et al.*, 2014; Jann, 2000; Gottwald; Rozanski, 1999). Foi desenvolvida em 1985 pela neurofarmacologista israelense Marta Weinstock-Rosin, durante seus estudos na Universidade Hebraica de Jerusalém, sendo posteriormente licenciada para a farmacêutica Novartis (Fischer; Ganellin, 2010).

Estudos bioquímicos demonstram que a rivastigmina inibe preferencialmente a hAChE no SNC, contrastando com sua menor ação em tecidos periféricos (Eldufani; Blaise, 2019; Gottwald; Rozanski, 1999; Polinsky, 1998). Sua ação dual, inibindo tanto a hAChE quanto a BChE, diferencia-se dos demais AChEIs, mais seletivos para a hAChE, conferindo à rivastigmina maior eficácia terapêutica. Essa característica contribui para a redução da formação de proteínas amiloides patogênicas, associadas à progressão da DA (Marucci *et al.*, 2021; Ray *et al.*, 2020).

Quimicamente, a rivastigmina possui um grupamento carbamato característico, com o nitrogênio ligado a um grupo etila e uma hidroxila funcional conectada ao restante da molécula. Os efeitos adversos mais comuns associados à rivastigmina são gastrointestinais, incluindo náuseas, vômitos e, com menor frequência, diarreia (Birks; Evans, 2015). Para minimizar esses efeitos, uma forma comum de administração é por via transdérmica.

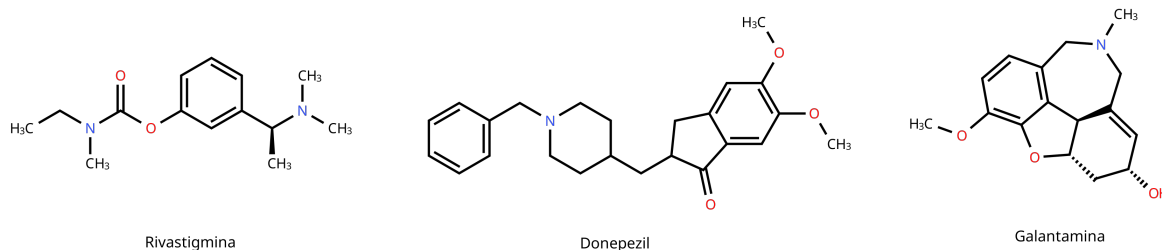
Diferentemente do donepezil e da galantamina, a rivastigmina apresenta atividade específica sobre a isoforma G₁ da hAChE, que está predominantemente localizada no córtex, no hipocampo e em placas neuríticas — regiões mais comprometidas pela deficiência colinérgica na DA (Vecchio *et al.*, 2021).

A galantamina é um inibidor reversível e seletivo da hAChE, aprovada pela FDA em 2001 para o tratamento sintomático da DA em estágios leve a moderado. Além de sua ação inibitória sobre a

hAChE, a galantamina também atua como modulador alostérico positivo dos receptores nicotínicos, potencializando a liberação de acetilcolina e promovendo efeitos neuroprotetores adicionais.

Sua estrutura química é composta por anéis fusionados de benzofurano e benzazepina, o que contribui para sua estabilidade e atividade biológica. Os efeitos adversos mais frequentemente relatados incluem náusea, vômito, diarreia, dor abdominal, astenia e insônia (Wilkinson *et al.*, 2001).

Figura 1.3: Representação em 2D das estruturas dos AChEIs clássicos



1.2.2 Compostos de partida: moléculas naturais e tacrina

Na presente seção, serão discutidos os compostos de partida, moléculas utilizadas no desenvolvimento de novos AChEIs. Essas moléculas foram selecionadas por apresentarem atividade em modelos da doença *in vitro* contra a enzima hAChE, e também por sua semelhança estrutural com inibidores clássicos da hAChE, como o DNP ou mesmo a THA, que, embora não seja mais utilizada como fármaco no tratamento da DA, ainda é amplamente empregada como composto de partida no desenvolvimento de novos AChEIs.

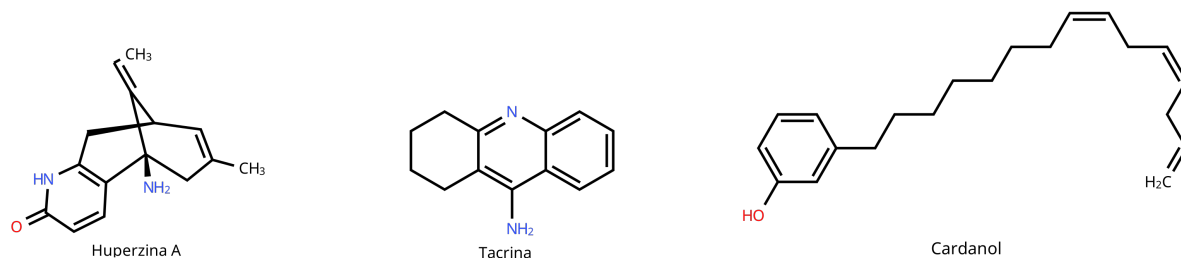
A maioria dessas moléculas é de origem natural, como a huperzina A, que é um sesquiterpeno alcaloide natural encontrado principalmente na planta *Huperzia serrata*, mas também presente em outras espécies do gênero, como *H. elmeri*, *H. carinat* e *H. aqualupian*. Amplamente utilizada na medicina tradicional chinesa, a huperzina A atua como inibidor da enzima hAChE e antagonista não competitivo do receptor NMDAR. Na China, é empregada no tratamento da doença de Alzheimer (Mak *et al.*, 2021), enquanto nos Estados Unidos está disponível como nutracêutico (Yang *et al.*, 2013). Embora estudos sugiram efeitos benéficos na melhora da função cognitiva e na avaliação clínica global em participantes de ensaios clínicos randomizados, é crucial considerar esses resultados com cautela devido à baixa qualidade metodológica das pesquisas analisadas (Yang *et al.*, 2013).

A fisostigmina é um alcaloide potente extraído do feijão-calabar (também conhecido como feijão-calvário), uma planta originária das florestas tropicais da África (Fraser, 1867). Apesar de sua capacidade de modular os receptores da colinesterase (ChE), a fisostigmina nunca foi amplamente utilizada clinicamente devido à sua baixa meia-vida e a uma série de efeitos adversos significativos (Coelho; Birks, 2001).

O cardanol é um lipídio fenólico não-isoprenoide, extraído das cascas da espécie *Anacardium occidentale* (Lim, 2011). Tem sido amplamente utilizado como composto de partida para o design de inibidores da acetilcolinesterase (AChEIs) do tipo *dual binding*. Os AChEIs *dual binding* são especialmente interessantes porque interagem simultaneamente com duas regiões importantes da enzima: o sítio ativo catalítico (CAS) e o sítio periférico aniônico (PAS) (Andrade Ramos *et al.*, 2021; Abreu Silva *et al.*, 2019; Lemes *et al.*, 2016).

Juntamente com a tacrina, compostos naturais como o cardanol são frequentemente empregados como base para o desenvolvimento de novos AChEIs na pesquisa farmacêutica.

Figura 1.4: Representação 2D das estruturas moleculares de compostos comumente usadas como partida para obtenção de novos AChEI.



1.2.3 AChEI/BChEI Multi-target

Após o desenvolvimento inicial dos AChEI/BChEI *single-target* e, com o auxílio dos compostos de partida discutidos na seção anterior, a pesquisa passou a se concentrar no desenvolvimento de novos fármacos para o tratamento da DA que fossem não apenas mais eficazes na interação com a hAChE, mas que também apresentassem a capacidade de atuar em outros alvos da doença, como as proteínas A β amiloides e as tauopatias, considerando a natureza multifatorial da doença.

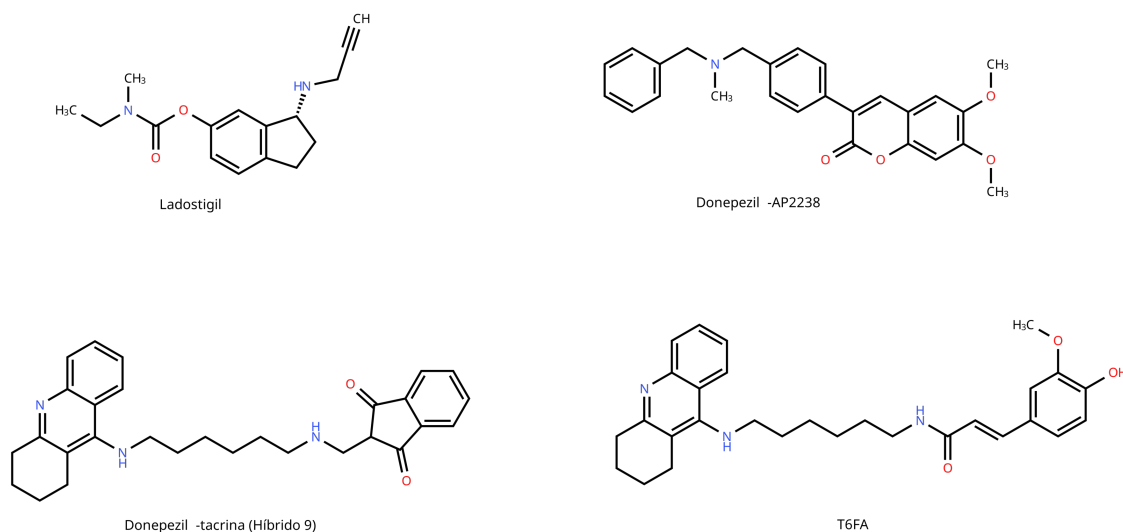
O primeiro candidato a fármaco com esse perfil a ser desenvolvido com sucesso foi o ladostigil, que demonstrou atividade promissora e, atualmente, encontra-se em estudo como agente neuroprotetor para o tratamento de doenças neurodegenerativas, como a DA, a DP e a Doença de corpos de Lewy (Shirbhate *et al.*, 2022; Wu *et al.*, 2022). Originado da modificação estrutural da rasagilina, já empregada no tratamento da DP, o ladostigil se destaca por seus mecanismos de ação multifuncionais. Ele atua como inibidor reversível das enzimas hAChE e BChE, além de ser um inibidor irreversível da monoamina oxidase B (MAO-B), combinando, assim, propriedades da rivastigmina e da rasagilina (Weinreb *et al.*, 2009; Weinstock *et al.*, 2006).

Devido à sua atividade *multi-target*, o ladostigil tem mostrado resultados promissores em estudos clínicos (Weinreb *et al.*, 2011; Schneider *et al.*, 2019). Apesar de sua atividade múltipla, o ladostigil não foi originalmente desenvolvido com essa finalidade específica. Dados os bons resultados obtidos com compostos que atuam em mais de um alvo para a DA, o desenvolvimento de moléculas híbridas com essa finalidade tem sido intensificado. Esses ligantes são geralmente projetados para se ligarem simultaneamente, assim como os derivados do cardanol, aos sítios CAS e PAS da hAChE.

A molécula donepezil-AP2238, um derivado do DNP foi o primeiro desses compostos a ser desenvolvido, demonstrando capacidade de se ligar a ambos os sítios da hAChE. Estudos indicaram que o donepezil-AP2238, com sua atividade dual na enzima hAChE, também exerce atividade antiagregação das proteínas A β amiloides. Essa característica contribui para um efeito multi-alvo da molécula, ampliando seu espectro de atividade na DA quando comparado ao donepezil isoladamente (Piazzi *et al.*, 2003; Rizzo *et al.*, 2010).

Outros compostos híbridos, como a tacrina conjugada com o ácido ferúlico (T6FA), mostraram-se mais eficazes do que a tacrina na ligação à hAChE *in vitro*, apresentando ainda menor toxicidade. O híbrido donepezil-tacrina também foi desenvolvido e demonstrou atividade inibitória nas enzimas hAChE, BChE e na A β (Mishra *et al.*, 2017; Alonso *et al.*, 2005).

Figura 1.5: Representação em 2D da estrutura geométrica de alguns AChEI/BChEI Multi-targets

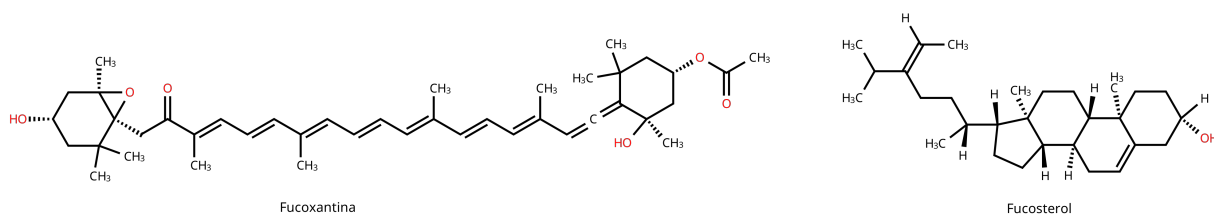


1.2.4 Metabolitos de algas

Alguns metabolitos isolados de algas têm demonstrado atividade inibitória sobre as enzimas ChE, com destaque para a fucoxantina. Esse carotenóide, extraído da classe de algas *Phaeophyceae*, liga-se de forma não competitiva ao PAS da hAChE, apresentando um $IC_{50} = 81.2\mu M$ em estudo *in vitro* com a hAChE (Lin *et al.*, 2016).

Outra molécula com potencial para a DA, obtida de diversas algas, é o fucosterol. Estudos indicaram que o fucosterol levou à redução da neurotoxicidade induzida pela proteína A β amiloide em neurônios de ratos localizados no hipocampo (Oh *et al.*, 2018).

Figura 1.6: Representação em 2D das estruturas dos compostos extraídos de algas com potencial atividade inibitória para as enzimas ChE.



1.2.5 Antagonistas do receptor NMDA

A memantina, um antagonista não competitivo dos NMDARs, teve uma origem inesperada, tendo sido inicialmente desenvolvida como um antidiabético pela Eli Lilly and Company em 1968 (Jack; Eriks, 1968). Sua atividade neuroprotetora foi descoberta por acaso durante estudos com a amantadina, seu composto precursor. Observou-se que pacientes com DP apresentavam uma redução significativa dos sintomas (Parsons *et al.*, 1999). Essa descoberta antecipou em anos a identificação dos NMDAR como o principal alvo biológico da memantina.

Os NMDAR são um dos três tipos de receptores de glutamato ionotrópicos, atuando como canais iônicos de cálcio nos neurônios, juntamente com os receptores AMPA (ácido α -amino-3-hidroxi-5-metil-4-isoxazol-propionato) e cainato. Na DA, a memantina age inibindo o influxo excessivo de íons cálcio gerado pela hiperestimulação dos NMDAR. Dessa forma, a memantina pode ser utilizada tanto em monoterapia quanto em combinação com AChEIs para o tratamento de diversas desordens neurológicas.

Tendo em vista a necessidade do desenvolvimento de novos fármacos — uma vez que os atualmente disponíveis atuam apenas no controle dos sintomas da doença e apresentam eficácia reduzida à medida que a progressão irreversível da DA se estabelece —, diversos estudos têm demonstrado o potencial terapêutico dos canabinóides, dos chamados canabinóides raros (*minor cannabinoids*) e de moléculas de origem natural conhecidas como canabinomiméticos, que emulam a atividade dos canabinóides, no tratamento da DA e de outras doenças neurodegenerativas (Coles *et al.*, 2022; Stone *et al.*, 2020; Doumar *et al.*, 2025).

Desta forma, na próxima seção, serão explorados os canabinóides como potenciais fontes para o desenvolvimento de novos AChEIs, com foco na sua aplicação terapêutica em doenças neurodegenerativas.

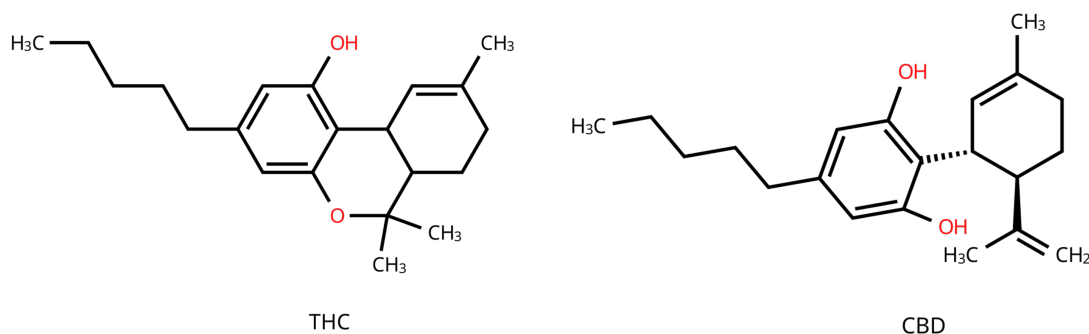
1.3 Canabinóides e canabinóides raros como potenciais agentes terapêuticos em doenças neurodegenerativas

Embora os AChEIs atuem no alívio sintomático de doenças neurodegenerativas como DA, DP e esclerose múltipla, apresentam eficácia limitada em impedir a progressão dessas doenças (Montero *et al.*, 2021). Esse cenário tem estimulado a busca por novas alternativas terapêuticas que sejam mais eficazes no alívio dos sintomas, capazes de impedir a progressão da doença ou que atuem de forma complementar aos tratamentos atualmente disponíveis.

Nesse contexto, os canabinóides — compostos derivados da planta *Cannabis sativa* —, bem como o sistema endocanabinóide (SEC), têm emergido como potenciais fármacos e alvos terapêuticos, respectivamente, para o tratamento de doenças neurodegenerativas. O SEC é composto por receptores canabinóides (principalmente o CB₁, presente no sistema nervoso central, e o CB₂, no sistema imune e no sistema nervoso periférico) e por ligantes endógenos, denominados endocanabinóides, como a anandamida (AEA) e o 2-arquidonilglicerol (2-AG). O SEC está amplamente distribuído pelo corpo humano, desempenhando papel em diversas funções críticas, como regulação do desenvolvimento cerebral, liberação de citocinas por micróglias, comportamento emocional e cognição (Galve-Roperh *et al.*, 2009; Kasatkina *et al.*, 2021; Zanettini *et al.*, 2011). Alterações no SEC estão associadas ao desenvolvimento de diversas patologias, incluindo doenças neurodegenerativas, o que torna a modulação farmacológica desse sistema um campo de grande interesse na pesquisa científica (Lowe *et al.*, 2021).

Os canabinóides naturais (fitocanabinóides) mais amplamente estudados são o Δ^9 -tetraidrocannabinol (THC) e o canabidiol (CBD) (Maroon; Bost, 2018). Embora o THC seja o principal componente psicoativo produzido pelas espécies de *Cannabis*, seu efeito psicotrópico — mediado principalmente pelos receptores CB₁ — limita seu uso terapêutico em larga escala. No entanto, o THC exerce efeitos neuroprotetores por meio da ativação dos receptores CB₁ e CB₂, atenuando a excitotoxicidade, reduzindo a atividade glial e eventos neuroinflamatórios locais (Sánchez; García-Merino, 2012; Gilbert *et al.*, 2007; Henriquez *et al.*, 2020). No contexto da DA, o THC demonstrou potencial terapêutico para o tratamento de aspectos comprometidos ao longo da doença, como memória, cognição e sintomas comportamentais (Tyrakis *et al.*, 2024).

Figura 1.7: Representação em 2D das estruturas dos dois principais canabinóides encontrados na *Cannabis Sativa L.*



O THC também apresentou ação na modulação das proteínas A β e na redução da fosforilação da proteína tau em modelos de DA (Cao *et al.*, 2014). No entanto, as evidências clínicas dessas propriedades ainda são limitadas (Fernández-Ruiz, 2019).

O CBD, por sua vez, não possui atividade psicoativa, pois se liga preferencialmente aos receptores CB₂ como antagonista. Atua também como agonista inverso dos receptores CB₁, o que justifica seu uso comum em associação com o THC, reduzindo os efeitos psicoativos deste último e potencializando os efeitos terapêuticos desejados nos receptores CB₂ (efeito entourage) (Peng *et al.*, 2022; Russo, 2011). Diversos estudos vêm demonstrando interesse no uso do CBD para o tratamento de doenças neurodegenerativas, diabetes e processos inflamatórios (Bhunia *et al.*, 2022; Jha *et al.*, 2024; Singh *et al.*, 2023; Zhang *et al.*, 2023; Rafailovska *et al.*, 2025; Catakli *et al.*, 2025; Hickey *et al.*, 2024).

Além de seus efeitos mediados pelo SEC, o CBD apresenta propriedades anti-inflamatórias e imunossupressoras, associadas ao aumento da sinalização de adenosina (Gonca; Darıcı, 2015). Também modula receptores de serotonina, especialmente o 5HT_{1A}, além de noradrenalina, dopamina e do ácido gama-aminobutírico (GABA), podendo inibir sua recaptação (Russo *et al.*, 2005; Cohen *et al.*, 2019), o que pode explicar suas propriedades ansiolíticas observadas (R de Mello Schier *et al.*, 2014).

Além do THC e do CBD, a *Cannabis sativa* contém mais de 500 compostos, sendo mais de 100 denominados fitocanabinóides já identificados, que são produzidas em quantidade menor na *Cannabis* e são comumente chamados de canabinóides raros (Rock; Parker, 2020; Walsh *et al.*, 2021). Existem ainda diversos compostos naturais que, embora não sejam derivados da planta *Cannabis*, apresentam propriedades químicas semelhantes aos dos canabinóides. Esses compostos são comumente denominados canabinomiméticos (Mevers *et al.*, 2014; Hudson *et al.*, 2010). Nas próximas seções, estes compostos, serão brevemente descritos.

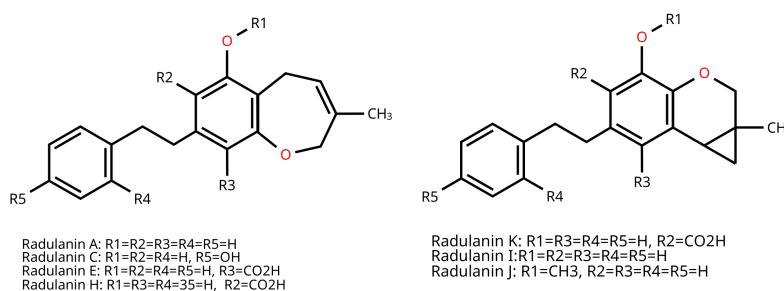
1.4 Gênero *Radula*

1.4.1 Radulaninas

Os compostos denominados radulaninas têm origem no gênero *Radula*, um grupo específico dentro da família *Radulaceae*, classificada na classe *Marchantiophyta*, comumente conhecida como hepáticas. Estima-se que existam cerca de 350 espécies dentro desse gênero em todo o mundo, sendo 60 destas espécies encontradas na Ásia (Yamada, 1979). O gênero *Radula* se distingue quimicamente de outras hepáticas, sendo predominantemente composto por bibenzilas, incluindo os chamados canabinóides bibenzílicos, derivados prenil-bibenzílicos e bis-bibenzilas (Asakawa *et al.*, 2020).

Diversos desses compostos demonstraram atividades biológicas variadas, como psicoatividade (Chicca *et al.*, 2018), ação antimicrobiana (Asakawa, 1990) e antifúngica (Lorimer *et al.*, 1993). As radulaninas são compostos químicos encontrados em várias espécies do gênero *Radula*. Esses compostos são bibenzilas que podem estar associados a di-hidrooxepina (por exemplo, radulaninas A, C, E e H), ou apresentar um anel cromeno modificado, fusionado com um anel ciclopropano, formando um sistema ciclopropa[c]cromeno (radulaninas I, K e J). São produzidos por diversas espécies, como *R. javanica*, *R. tokiensis*, *R. buccinifera*, entre outras.

Figura 1.8: Principais radulaninas



A radulanina A, radulanina C e a radulanina E são produzidas concomitantemente pelas espécies *R. buccinifera* (exceto o radulanina E), *R. complanata*, *R. japonica*, *R. javanica* e *R. tokiensis* (Asakawa *et al.*, 2020). Dessa forma, a radulanina A, por ser um metabólito comum em diversas espécies do gênero *Radula*, é utilizado como marcador na quimiosistemática de grupos específicos dentro desse gênero.

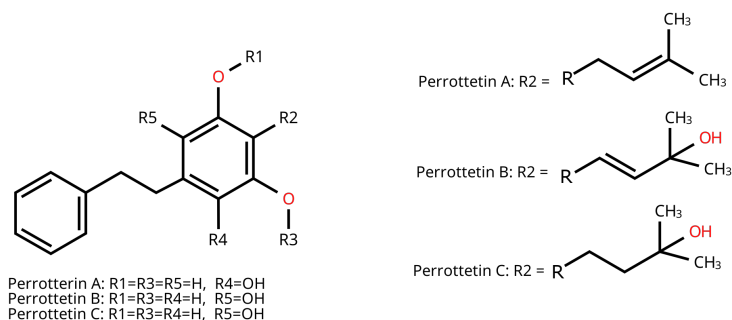
As radulaninas A, L e K, juntamente com o composto 2-geranil-3,5-di-hidroxibibenzil, foram isolados da espécie *R. appressa* e avaliados quanto à inibição do óxido nítrico (NO) em um estudo *in vitro* com células RAW 264.7, uma linhagem de macrófagos de camundongo (Harinantenaina *et al.*, 2006). A liberação de NO foi estimulada pela presença de lipopolissacarídeo (LPS), um componente importante da membrana externa de bactérias gram-negativas, conhecido por ser um potente estimulador da resposta do sistema imunológico (Choudhury *et al.*, 2018; Boncoeur *et al.*, 2013; Miyazaki *et al.*, 2012).

Todos os bibenzilas avaliados no estudo mostraram inibição da produção de NO, com concentrações inibitórias de 50% variando de 4,5 a 20 μ M, sendo que o radulanina L e o 2-geranil-3,5-di-hidroxibibenzil apresentaram os melhores resultados. Além disso, outro estudo demonstrou que o radulanina K inibiu a liberação do radical superóxido de macrófagos de camundongos, com um IC₅₀ de 6 μ g/mL (Schwartner *et al.*, 1996).

1.4.2 Canabinóides bibenzílicos

Os primeiros prenil-bibenzilas, denominados perrotetinenos A-D, apresentaram atividade antimicrobiana contra a bactéria *Streptococcus aureus*, tendo sido isolados da espécie *R. perrottetii* (Asakawa *et al.*, 1982). A estrutura básica dos prenil bibenzilas é um núcleo bibenzílico que consiste em dois anéis benzênicos conectados por dois átomos de carbono e pela presença de um ou mais grupos prenila, que são derivados do isopreno, ligados a estrutura bibenzílica.

Figura 1.9: Esquema estrutural dos principais perrotetinenos

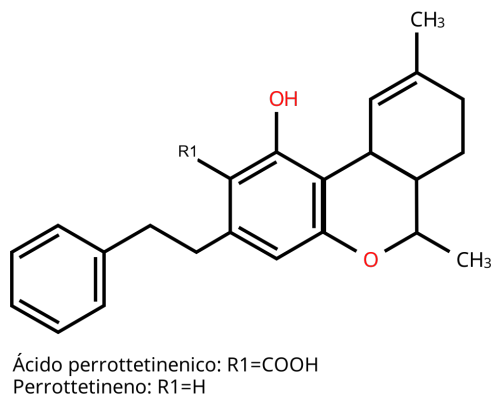


Todavia, dentre os prenil bibenzilas, destacam-se os canabinóides bibenzílicos, que viriam a ser identificados em 2002 na Nova Zelândia com o isolamento do perrotetino (PRT) e do ácido perrotetinênico da espécie *Radula marginata* (Toyota *et al.*, 2002). PRT foi também encontrado nas espécies típicas da Nova Zelândia, Japão e Costa Rica - *R. perrottetii* (Toyota *et al.*, 1994), *R. campanigera* e *R. chinensis* (Miyazawa; Asakawa, 2018) e *R. laxiramea* (Cullmann; Becker, 1999). O ácido perrotetinênico também foi isolado da hepática *Plagiochila ovalifolia*, pertencente a família Phagiochilaceae (Sadamori, 2009).

Os canabinóides bibenzílicos associam a estrutura dos prenil-bibenzilas com o anel benzo[c]cromênico encontrado nos canabinóides. A descoberta dos canabinóides bibenzílicos atraiu a atenção de diversos pesquisadores como químicos, farmacologistas e bioquímicos. Chicca e colaboradores publicaram um artigo de revisão sobre a psicoatividade desses compostos, neste mesmo artigo, os autores descreveram a síntese total estereoseletiva do *cis*-perrotetinenos e do *trans*-perrotetino e reportaram sobre a sua farmacologia *in vitro* e *in vivo* em detalhe (Chicca *et al.*, 2018).

Os dois diastereoisômeros do perrotetino penetraram a barreira hematoencefálica, e induziram hipotermia, catalepsia, hipocolocomotria e analgesia associada a atividade agonista semelhante a do THC, em receptores CB₁ em camundongos. Todavia, o isômero *cis* interagiu com um maior número de receptores CB₁ no cérebro, demonstrando maior seletividade farmacológica quanto ao sistema canabinóide.

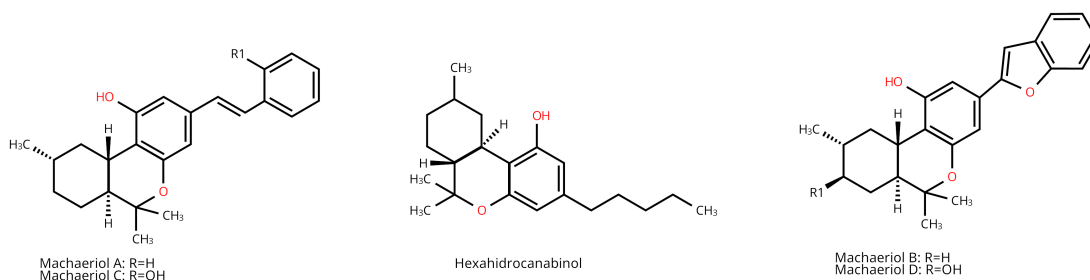
Figura 1.10: Esquema estrutural dos principais canabinóides bibenzílicos-perrottetino e ácido perrottetênico



1.5 Gênero *Machaerium*

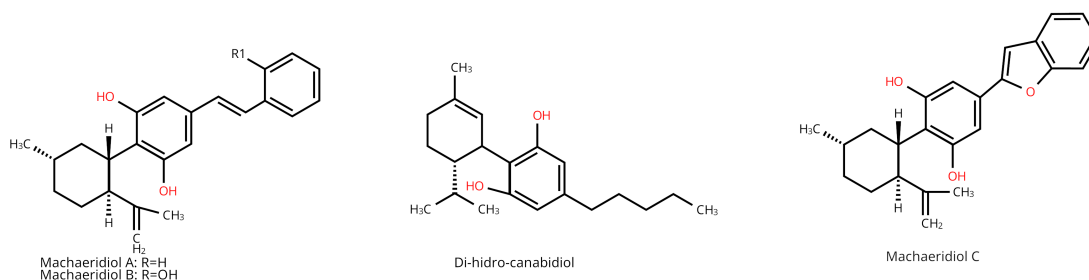
O gênero *Machaerium* contém aproximadamente 130 espécies distribuídas nas Américas tropicais, sendo encontradas em arbustos, lianas e em árvores de pequeno e médio porte (Mabberley, 2017; Amen *et al.*, 2015; Joly *et al.*, 1987; Polhill; Raven, 1981). Todavia os compostos dos machaeríols e machaerídíols foram encontrados apenas na espécie *Machaerium Pers* (Kumarihamy *et al.*, 2022), tendo sido isolados da casca do caule da planta e coletadas em solos arenosos de florestas abertas na província de Mayanas, perto da região de Loreto no Peru. Os machaeríols e machaerídíols são fitocanabinóides aralquílicos raros, tendo elevada semelhança estrutural com canabinóides clássicos. Os machaeríols A-D são fitocanabinóides do tipo hexahidrodibenzopirano (HHDBP), sendo análogos estruturais do hexahidrocanabinol (HHC) (Figura 1.11).

Figura 1.11: Representação 2D das estruturas químicas dos machaeríols com HHC no centro



Já os machaerídíols A-C são fitocanabinóides do tipo 5,6-seco-hexahidrodibenzopirano (seco-HHDBP), sendo análogos do di-hidrocanabidiol (Figura 1.12). Os machaeríols e machaerídíols demonstraram potente atividade antimicrobiana, especialmente contra cepas de *Staphylococcus aureus* resistente à metilicina (MRSA) e *Enterococcus faecium* e *Enterococcus faecalis* resistentes à vancomicina (VRE), tendo os machaerídíols A-C demonstrado maior potencia contra MRSA do que o canabidiol (CBD) e a vancomicina (Muhammad *et al.*, 2020). Tanto os machaeríols como os machaerídíols apresentaram atividade inibitória contra malária e leishmania (Muhammad *et al.*, 2001, 2003).

Figura 1.12: Representação 2D das estruturas químicas dos machaeridiols com di-hidrocanabidiol no centro



Não foi encontrada atividade nos receptores CB₁ para nenhum dos machaeridiols ou machaeridiols, indicando segurança relacionada a efeitos psicoativos associados a este receptor. O machaeriol D e os machaeridiols A-C, demonstraram relevante afinidade seletiva pelos receptores canabinóides CB₂, de forma semelhante as relatadas para o canabidiol. Tanto os machaeridiols como os machaeridiols obtiveram fraca atividade nos receptores opioides δ e μ (Muhammad *et al.*, 2023).

A citotoxicidade em células tumorais sólidas *in vitro* também foi avaliada no mesmo estudo. Alguns compostos apresentaram citotoxicidade moderada, como o machaeridiol B que teve atividades fracas contra células HeLa, SK-MEL, KB e BT-549, com valores de IC₅₀ entre 32 e 38 $\mu\text{g/mL}$. Uma combinação de machaeriol C com o machaeridiol C em proporção 1:1 demonstrou atividade aditiva fraca contra células HeLa, KB, BT-549 e SK-OV-3.

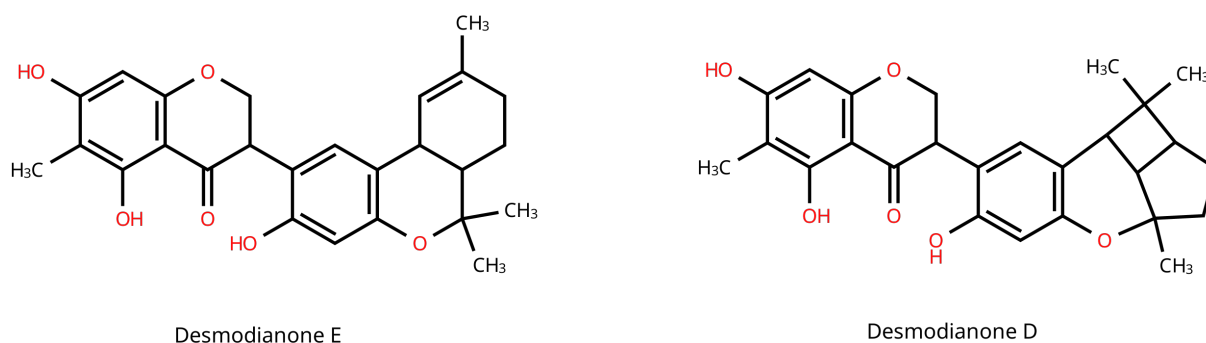
Os compostos machaeriol B e D merecem atenção em relação ao tratamento de DA, devido a presença do anel benzofurano (Figura 1.11), tendo em vista que derivados do benzofurano vem sendo estudado como potenciais AChEI (Goyal *et al.*, 2018; Cabrera-Pardo *et al.*, 2020; Abd El-Karim *et al.*, 2023).

1.6 Gênero *Desmodium*

A espécie *Desmodium incanum* é uma planta herbácea pertencente à família das leguminosas, conhecida por sua capacidade de fixação de nitrogênio, sendo seu estudo de interesse predominante nas áreas agroflorestal e medicinal (Toniutti *et al.*, 2017; Imrie *et al.*, 2019; Cueva-Chamba *et al.*, 2023). É encontrada na América, África e Ásia tropicais, sendo comum em regiões tropicais úmidas, crescendo em áreas abertas (Imrie *et al.*, 2019). A *Desmodium incanum* produz as principais desmodianonas, que são moléculas com potencial terapêutico diverso, encontradas na planta. As desmodianonas são isoflavonas preniladas, um tipo de flavonóide com cadeia lateral contendo grupos isoprenil (Simons *et al.*, 2012). As diferenças principais nestes compostos estão na posição e no tipo de substituintes aromáticos e nas cadeias laterais (Figura 1.13), o que influencia suas propriedades biológicas.

Estudos iniciais destas moléculas indicam potencial farmacológico em diversas áreas, como na atividade antimicrobiana, estrogênica, propriedades antioxidantes e potencial antitumoral (Lattibeaudiere; Newman, 2024).

Figura 1.13: Representação 2D das desmodianonas usadas no presente trabalho.

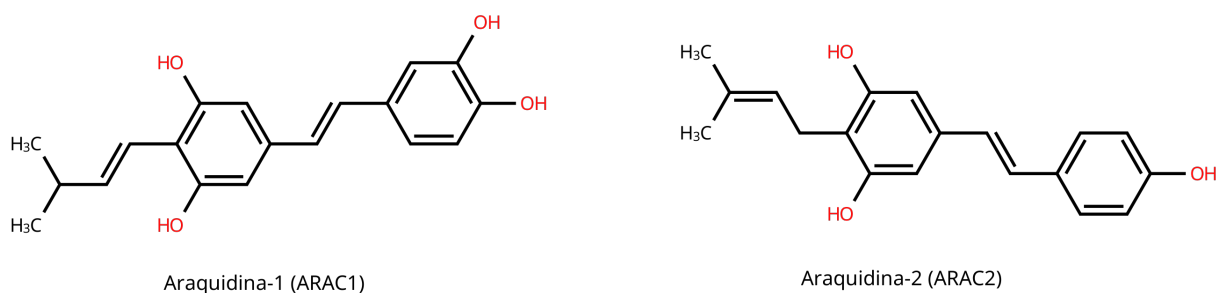


1.7 Gênero *Arachis*

As araquidinas são uma classe de estilbenos prenilados, encontrados principalmente no amendoim (*Arachis hypogaea*), bem como em seus ancestrais selvagens *A. duranensis* e *A. ipaensis* (Fang *et al.*, 2020b). Os estilbenos prenilados bem como o resveratrol produzidos no amendoim são comumente denominados fitoalexinas, que são compostos químicos com atividade microbiana produzidos por determinada planta em resposta a infecções ou ao estresse (Hasan *et al.*, 2013).

As araquidinas são caracterizadas por uma estrutura C6-C2-C6 (1,2-difeniletileno), com dois anéis aromáticos conectados por uma ponte de etileno. A araquidina-1 (ARAC1) possui uma porção 3-metil-1-butenila, enquanto a araquidina-2 (ARAC2) possui uma porção 3,3-dimetilalila (Figura 1.14). Essa porção prenilada nas araquidinas é crucial para sua atividade biológica.

Figura 1.14: Representação 2D das principais araquidinas encontradas na *Arachis hypogaea* e estudadas no presente trabalho.



As araquidinas e os demais estilbenos tem sido pesquisados em estudos pré-clínicos referentes a sua atividade contra doenças neurodegenerativas como a DA, tendo os estudos demonstrado efeitos neuroprotetores, influenciando a redução da geração e da oligomerização das proteínas A β amiloides, diminuindo as espécies reativas de oxigênio e consequentemente o estresse oxidativo, e prevenindo a fosforilação e a agregação anormal da proteína tau (Li *et al.*, 2018; Chen *et al.*, 2019; Al-Edresi *et al.*, 2020; Mahmoud Moustafa *et al.*, 2021; Chen *et al.*, 2018).

Capítulo 2

Objetivos

2.1 Objetivo

Avaliar o potencial de canabinóides e canabinomiméticos como AChEIs por meio de ferramentas de docking molecular, cálculos de estrutura eletrônica e análise de descritores de propriedades farmacocinéticas (ADMET) e de reatividade molecular.

2.2 Objetivos Específicos

- Identificar e analisar os principais descritores ADMET e de estrutura eletrônica relevantes para o sistema em estudo;
- Desenvolver e implementar algoritmos de machine learning voltados para a análise e interpretação dos dados obtidos;
- Construir modelos baseados em redes neurais para estudos de similaridade molecular, visando à triagem e ao desenho de novas moléculas com potencial atividade inibitória frente à hAChE.

Capítulo 3

Referencial Teórico

3.1 Teoria do Funcional de Densidade- DFT

A teoria DFT é utilizada para calcular a estrutura eletrônica de átomos e moléculas, tendo sido desenvolvida por Hohenberg e Kohn em 1964 (Hohenberg; Kohn, 1964). A teoria de funcional de densidade (do inglês *density functional theory*, DFT) define a energia do estado fundamental como um funcional da densidade eletrônica ($E[\rho]$), calculado a partir da soma dos demais funcionais de densidade do sistema

$$E[\rho] = T_{KS}[\rho] + J[\rho] + E_{xc}[\rho] + V_{ext}[\rho], \quad \rho(\mathbf{r}) = \sum_{i=1}^N |\psi_i|^2 \quad (3.1)$$

onde ψ_i é a função de onda monoelétrica, $\rho(\mathbf{r})$ a densidade eletrônica, $T_{KS}[\rho]$ o funcional correspondente a energia cinética,

$$T_{KS}[\rho] = -\frac{1}{2} \sum_{i=1}^N \langle \psi_i(\mathbf{r}) | \nabla_{\mathbf{r}}^2 | \psi_i(\mathbf{r}) \rangle \quad (3.2)$$

$J[\rho]$ o potencial de Coulumb de interação eletrônica, sendo definido como

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_2 - \mathbf{r}_1|} d\mathbf{r}_2 d\mathbf{r}_1 \quad (3.3)$$

$E_{xc}[\rho]$ o funcional de troca e correlação, cuja expressão exata não é conhecida, o funcional de troca e correlação é o principal alvo de desenvolvimentos da DFT. Por último o funcional $V_{ext}[\rho]$ está relacionado a energia potencial externa gerada pelos núcleos do sistema,

$$V_{ext}[\rho] = - \sum_{n=1}^{N_{\text{núcleo}}} \int Z_n \frac{\rho(\mathbf{r})}{|\mathbf{R}_n - \mathbf{r}|} d\mathbf{r} \quad (3.4)$$

sendo Z_n a carga nuclear do n -ésimo núcleo do sistema, $\mathbf{R}_n$ o vetor posição com as coordenadas do núcleo e $\mathbf{r}$ o vetor posição associado a densidade eletrônica ρ .

3.2 Índices de Reatividade Química derivados da teoria do funcional de densidade

A partir da densidade eletrônica $\rho(\mathbf{r})$ obtida dos cálculos da teoria do funcional de densidade discutida na seção anterior, é possível obter quantitativamente, medidas químicas usadas na classificação da reatividade química, sendo esta derivação chamada de DFT conceitual-CDFT (Geerlings *et al.*, 2003).

A primeira propriedade de reatividade química que pode ser obtida via DFT é o potencial químico, μ , que é igual ao multiplicador de Lagrange usado na equação de densidade eletrônica de Kohn-Sham (Parr *et al.*, 1978). O potencial químico dos elétrons do sistema, mede a tendência do fluxo de elétrons se dirigir para fora do sistema químico em equilíbrio (Atkins *et al.*, 2023), sendo definido como a derivada parcial da energia do sistema em relação ao número de elétron, mantendo o potencial externo, $V(\mathbf{r})$ fixo.

$$\mu = \left(\frac{\partial E}{\partial N(\mathbf{r})} \right)_{V(\mathbf{r})}. \quad (3.5)$$

Outra importante propriedade extraída da teoria do funcional de densidade é a função de Fukui. A função de Fukui descreve a densidade eletrônica associada a remoção ou adição de elétrons no sistema, sendo utilizada para modelar quantitativamente, os valores para o ataque nucleofílico, eletrofílico e radicalar.

$$f(\mathbf{r}) = \left(\frac{\partial \mu}{\partial V(\mathbf{r})} \right)_N = \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_{V(\mathbf{r})}. \quad (3.6)$$

Parr e Yang propuseram três definições numéricas baseadas na função de Fukui, que são comumente denominadas índices de Fukui ($f^+(\mathbf{r})$, $f^-(\mathbf{r})$ e $f^0(\mathbf{r})$, representando, respectivamente, a tendência para ataques nucleofílico, eletrofílico e radicalar). Essas definições auxiliam na identificação da reatividade das moléculas, oferecendo uma alta precisão quando comparadas aos dados experimentais (Domingo *et al.*, 2016).

$$\begin{aligned} f^+(\mathbf{r}) &= \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_{V(\mathbf{r})}^+ \approx \rho_{N_0}(\mathbf{r}) - \rho_{N_0-1} = \rho_{\text{HOMO}}(\mathbf{r}) \\ f^-(\mathbf{r}) &= \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_{V(\mathbf{r})}^- \approx \rho_{N_0+1}(\mathbf{r}) - \rho_{N_0} = \rho_{\text{LUMO}}(\mathbf{r}) \\ f^0(\mathbf{r}) &= \left(\frac{\partial \rho(\mathbf{r})}{\partial N} \right)_{V(\mathbf{r})}^0 \approx \frac{1}{2} (f^+(\mathbf{r}) + f^-(\mathbf{r})) \end{aligned} \quad (3.7)$$

sendo ρ_N a densidade eletrônica em um ponto $\mathbf{r}$ no espaço ao redor da molécula neutra. O sistema $N + 1$ representa o ânion, obtido pela adição de um elétron ao orbital LUMO da molécula neutra. Enquanto que, o sistema $N - 1$ corresponde ao cátion, resultante da remoção de um elétron do orbital HOMO. Em todas as três situações (N , $N + 1$ e $N - 1$), a densidade eletrônica é calculada mantendo-se a geometria da molécula neutra.

O descritor duplo condensado, CDD (do inglês *condensed dual descriptor*), é um índice importante derivado da teoria DFT, que permite identificar, em nível atômico, quais regiões de uma molécula apresentam maior tendência a se comportar como nucleofílicas ou eletrofílicas. Dado o átomo k de um sistema molecular, o CDD é definido como

$$\text{CDD}_k = f_k^+ - f_k^-. \quad (3.8)$$

No presente trabalho, utilizamos as notações CDD^+ e CDD^- para indicar, respectivamente, o maior e o menor valor de CDD. Parâmetros adicionais utilizados para descrever a reatividade química, como o gap HOMO–LUMO, o índice de eletrofilicidade e a eletronegatividade, são determinados com base nas energias dos orbitais de fronteira (HOMO e LUMO) e no potencial químico.

3.3 Interações não-covalentes

As interações não covalentes (*non-covalent interactions* – NCI) correspondem a forças inter e intramoleculares que atuam entre íons, átomos ou moléculas, sem envolver o compartilhamento direto de pares de elétrons, como nas ligações covalentes. Apesar de, em geral, apresentarem intensidade significativamente menor que as interações covalentes, são essenciais para a estabilidade, organização estrutural e funcionalidade de uma ampla variedade de sistemas químicos e bioquímicos.

As NCI podem ser classificadas com base na distância entre os átomos envolvidos, no tipo de espécies químicas participantes e na geometria da interação. Dentre essas, destacam-se as ligações de hidrogênio, que ocorrem entre um átomo de hidrogênio ligado covalentemente a um átomo eletronegativo (como oxigênio, nitrogênio ou flúor) e outro átomo eletronegativo adjacente. Essas ligações apresentam distâncias médias entre 1,5 – 2,2 Å (1 Å = 1×10^{-10} m). São essenciais para a estrutura de biomoléculas como o DNA e proteínas, além de desempenharem papel crucial na solvatação em ambientes aquosos, sendo fundamentais para as ligações peptídicas (Desiraju; Steiner, 2001; Fernández, 2020; Pattanayak; Chowdhuri, 2011).

As interações íon-dipolo e dipolo-dipolo envolvem, respectivamente, a atração entre íons e moléculas polares, ou entre moléculas polares entre si. Essas interações são relevantes em processos de solubilização, catálise e reconhecimento molecular (Hughes *et al.*, 2021; Adams *et al.*, 2025; Daze; Hof, 2016). As distâncias típicas variam de 2,0 – 3,5 Å para íon-dipolo e de 2,0 – 4,0 Å para dipolo-dipolo (Warshel, 1978; Buckingham, 1967).

As forças de van der Waals, também chamadas de forças de dispersão de London, surgem de flutuações temporárias na distribuição eletrônica de moléculas. Apesar de sua natureza fraca, são fundamentais para a coesão de sistemas apolares, como em membranas lipídicas, e contribuem significativamente para

a estrutura terciária de proteínas (Lee, 2003; Rehman *et al.*, 2024). As distâncias médias dessas interações situam-se entre 3,0 – 4,0 Å (Israelachvili, 2011).

As interações pi-pi e cátion-pi ocorrem frequentemente em sistemas contendo anéis aromáticos. São particularmente relevantes em química supramolecular, na cristalografia e em mecanismos de interação entre fármacos e seus alvos biológicos (Keiluweit; Kleber, 2009; Desiraju *et al.*, 2011; Ziada *et al.*, 2022). As distâncias típicas variam de 3,3 a 3,8 Å para interações pi-pi e de 2,5 a 3,5 Å para interações cátion-pi (Hunter; Sanders, 1990; Ma; Dougherty, 1997).

Por fim, as interações hidrofóbicas ocorrem predominantemente entre grupos apolares ou cadeias laterais não polares, sendo cruciais para o enovelamento de proteínas, a formação de membranas biológicas e a auto-organização molecular (Nelson; Cox, 2022). As distâncias associadas a essas interações geralmente variam entre 3,5 – 5,0 Å (Chandler, 2005).

3.4 Gradiente de densidade reduzida-RDG

O gradiente de densidade reduzida $s(\mathbf{r})$ (do inglês *reduced density gradient*, RDG) é uma função escalar que quantifica a variação espacial da densidade eletrônica $\rho(\mathbf{r})$ em torno de um ponto no espaço.

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{2(3\pi^2)^{\frac{1}{3}}\rho(\mathbf{r})^{\frac{4}{3}}} \quad (3.9)$$

onde $\nabla\rho(\mathbf{r})$ é o gradiente de densidade eletrônica, indicando a mudança de densidade ao longo do espaço, $\rho(\mathbf{r})$ é a densidade eletrônica no ponto $\mathbf{r}$ e o denominador da expressão é uma normalização baseada na densidade homogênea do gás de elétrons, típica no DFT (Perdew *et al.*, 1996).

Valores baixos de $s(\mathbf{r})$ ocorrem em regiões onde a densidade eletrônica varia de forma suave, como no interior dos átomos ou ao longo de ligações covalentes, enquanto valores altos indicam variações bruscas na densidade, típicas de regiões entre átomos que não compartilham ligações covalentes. Esse comportamento é explorado no método desenvolvido por Johnson e colaboradores, cujo objetivo é visualizar as regiões de interações não covalentes em uma molécula (Johnson *et al.*, 2010).

Para isso, utiliza-se o chamado diagrama RDG, um gráfico que representa o gradiente de densidade reduzida $s(\mathbf{r})$ no eixo y e no eixo x o produto entre o sinal de λ_2 com $\rho(\mathbf{r})$, $sign(\lambda_2)\rho(\mathbf{r})$, onde λ_2 é o segundo autovalor da matriz Hessiana da densidade eletrônica, $\rho(\mathbf{r})$.

Esse gráfico permite distinguir regiões de interações covalentes, associadas a picos negativos do produto $\lambda_2\rho(\mathbf{r})$, regiões de interações não covalentes fracas, onde $s(\mathbf{r})$ apresente valores próximos de zero, e regiões de repulsão estérica, indicadas por valores positivos de $\lambda_2\rho(\mathbf{r})$. Dessa forma, os diagramas RDG consolidam-se como uma ferramenta fundamental para a visualização e interpretação das interações não covalentes, promovendo uma ponte conceitual entre a análise baseada em densidade eletrônica e os princípios do CDFT.

3.5 Docking Molecular

O docking molecular é uma técnica computacional com ampla utilização na modelagem molecular, tendo como objetivo prever a orientação mais favorável de um ligante, ao interagir com um receptor biológico, como proteínas e enzimas. O propósito central desta técnica é estimar o modo de ligação (pose) e a afinidade entre ligante-alvo biológico, através de uma função de pontuação (score), comumente chamada de energia de ligação ou de interação.

Portanto, o docking molecular é uma ferramenta essencial no desenvolvimento moderno de novos fármacos. Historicamente, a consolidação dos estudos de docking se deu nas décadas de 1980 e 1990, com avanços significativos nas áreas da química e da biologia computacional. Um exemplo importante, foi a criação do software AutoDock, pelos cientistas do *The Scripps Research Institute*, Garret M. Morris e David S. Goodsell (Morris *et al.*, 2001). O algoritmo do AutoDock utiliza algoritmos genéticos para explorar o espaço conformacional dos ligantes, utilizando uma função de energia semiempírica para estimar a energia de ligação.

Posteriormente, foi desenvolvido o algoritmo AutoDock Vina, pelos pesquisadores Oleg Trott e Arthut J. Olsen, também do Scripps, fornecendo um avanço considerável em desempenho e precisão (Trott; Olson, 2010). O algoritmo Vina introduz uma nova função de pontuação híbrida, que combina componentes físicos e estatísticos do sistema, além de ter a adição do suporte a processamento *multithreading*, tornando-o consideravelmente mais rápido do que seu antecessor.

A função de pontuação em algoritmos de docking molecular estima a energia livre de ligação, ΔG entre o ligante e o receptor, sendo calculada a partir da seguinte expressão

$$\Delta G = RT \ln K_i \approx \Delta G_{\text{ligação}} = \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} + \Delta G_{\text{dess}} + \Delta G_{\text{tors}} + \Delta G_{\text{H-bond}} \quad (3.10)$$

em que os termos do lado direito da expressão são, respectivamente, as contribuições das interações de van der Waals, eletrostáticas, de dessolvatação, de torsão e interação de hidrogênio. Cada algoritmo de docking molecular obtém estes elementos da equação da função de pontuação de forma própria, utilizando aproximações experimentais e métricas estatísticas.

Apesar das diferenças, tanto AutoDock como o Vina são amplamente utilizados na pesquisa científica atual. O AutoDock geralmente é preferido em contextos que exigem o controle fino sobre os parâmetros físicos da simulação, enquanto o Vina é comumente usado em pipelines automatizadas e em estudos de alto desempenho, com grande volume de dados, devido a à sua rapidez e simplicidade de uso (Morris *et al.*, 2009; Forli *et al.*, 2016).

A validação dos estudos de docking molecular é uma etapa fundamental para garantir a confiabilidade dos resultados obtidos com as energias de ligação preditas. Nesse contexto, destaca-se o método denominado redocking, que consiste em comparar a pose prevista pelo protocolo de docking utilizado, com a pose experimental do ligante complexado na enzima ou proteína, obtida pelo complexo cristalizado da estrutura disponível em bancos de dados de proteínas, como o PDB.

Esse comparação é feita utilizando a métrica RMSD (do inglês *Root Mean Square Deviation*), que quantifica o desvio quadrático médio entre as posições atômicas previstas pelo protocolo, com as posições reais do ligante no cristal

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\vec{r}_i^{\text{dock}} - \vec{r}_i^{\text{ref}})^2} \quad (3.11)$$

onde N é o número total de átomos considerados, $\vec{r}_i^{\text{dock}}$ são as coordenadas dos átomos na pose prevista, $\vec{r}_i^{\text{ref}}$ as coordenadas dos átomos do ligante na estrutura cristalográfica. Valores de RMSD inferiores a 2.0 Å são comumente aceitos na literatura como indicativo de boa reprodutibilidade e da validação do protocolo de docking molecular utilizado.

Na próxima seção, serão apresentados os métodos de ciência de dados utilizados e investigados neste trabalho, os quais complementam e aprofundam a análise teórica das interações moleculares, com foco na descoberta e no reposicionamento de fármacos.

3.6 Métodos em Data Science

3.6.1 Gradiente descendente como método central de otimização em machine learning

O algoritmo do gradiente descendente (GD) é um algoritmo com ampla aplicação na ciência, podendo ser empregado em diversas áreas. Em métodos de ciência de dados, é comumente empregado como um algoritmo de regressão próprio de aprendizado de máquina, porém sua maior aplicação é como uma ferramenta utilizada por grande parte dos algoritmos de aprendizado de máquina e em arquiteturas de redes neurais, para minimização do erro durante a etapa de aprendizagem dos modelos. Dado determinado valor alvo (*target-value*), y , o algoritmo GD é inicializado com a chamada hipótese, $h_{\Theta}(x)$,

$$h_{\Theta}(x) = \sum_i \Theta_i x_i \quad (3.12)$$

onde Θ_i são os pesos atribuídos a cada valor de entrada x_i . O objetivo do algoritmo é minimizar a diferença entre a hipótese e o valor-alvo (*target-value*), o que pode ser feito através da função de custo, $J(\Theta)$, definida como o erro quadrático médio entre ambos. Essa função é derivada do quadrado da distância euclidiana escalar entre a hipótese, e o *target-value*, dada por

$$J(\Theta) = \frac{1}{2} (h_{\Theta}(x) - y)^2. \quad (3.13)$$

O algoritmo a cada iteração atualiza os pesos Θ_i , afim de minimizar a função de custo

$$\Theta_j := \Theta_j - \alpha \left(\frac{\partial J(\Theta)}{\partial \Theta_j} \right) \quad (3.14)$$

sendo α a taxa de aprendizado (learning rate), que é utilizada para calibrar a atualização do peso Θ_j a cada iteração, e o sinal “:=” representando que a cada iteração, o peso Θ_j recebe um novo valor (é atualizado). Portanto, para resolver esta expressão, é necessário obter o valor da derivada parcial da função de custo

$$\frac{\partial J(\Theta)}{\partial \Theta_j} = \frac{1}{2} \frac{\partial}{\partial \Theta_j} (h_{\Theta}(x) - y)^2 = \frac{1}{2} \frac{\partial}{\partial \Theta_j} \left(\sum_i \Theta_i x_i - y \right)^2 \quad (3.15)$$

derivando 3.15

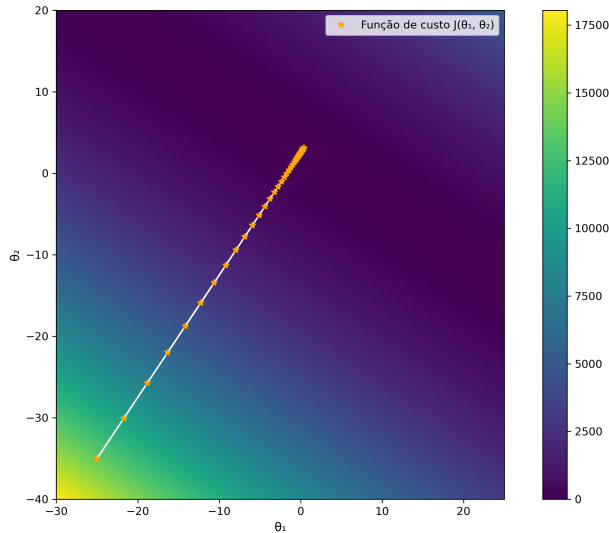
$$\frac{1}{2} \frac{\partial}{\partial \Theta_j} \left(\sum_i \Theta_i x_i - y \right)^2 = \left(\sum_i \Theta_i x_i - y \right) x_j = (h_{\Theta}(x) - y) x_j \quad (3.16)$$

e substituindo 3.16 na expressão 3.14

$$\Theta_j := \Theta_j + \alpha(y - h_{\Theta}(x)) x_j. \quad (3.17)$$

A Figura 3.1 ilustra uma simulação do algoritmo GD aplicado à minimização de uma função de custo quadrática oriunda de um problema de regressão supervisionada com dois parâmetros. Nesse contexto, a hipótese é definida como $h_{\Theta}(x) = \Theta_1 x_1 + \Theta_2 x_2$, com vetor de entrada $\mathbf{x} = [2, 0, 3, 0]$ e target-value $y = 10, 0$. A função de custo associada é dada por $J(\Theta_1, \Theta_2) = \frac{1}{2} (h_{\Theta}(x) - y)^2$, e o objetivo do algoritmo é encontrar os valores de Θ_1 e Θ_2 que minimizem o erro quadrático entre a predição e o target-value. O algoritmo é inicializado com valores arbitrários $\Theta_1 = -25$ e $\Theta_2 = -35$, correspondendo ao erro elevado inicial na predição. A cada iteração, os parâmetros são atualizados com base no gradiente da função de custo, reduzindo progressivamente o erro. No gráfico, a superfície colorida representa os valores de $J(\Theta_1, \Theta_2)$ para diferentes combinações de pesos, enquanto as estrelas marcam a trajetória do algoritmo em direção ao mínimo. A linha que conecta as estrelas demonstra o caminho percorrido pelo algoritmo até a convergência.

Figura 3.1: Exemplo de simulação da trajetória bidimensional do Algoritmo gradiente descendente



3.7 Algoritmos de classificação

3.7.1 Máquinas de vetores de suporte

Introdução

As máquinas de vetores de suporte (do inglês *support vector machines*, SVM), pertencem a um grupo de algoritmos de machine learning utilizados para classificação e regressão. Foi desenvolvido pelos cientistas Vladimir N. Vapnik e Alexey Ya. Chervonenkis na Bell Laboratories em 1964 (Vapnik; Chervonenkis, 1964). Na área da saúde, o SVM possui diversas aplicações como na classificação de tipos de tumores cerebrais (Vani *et al.*, 2017), detecção de tumores de mama (Chang *et al.*, 2003), modelagem QSAR (Mei *et al.*, 2005) e na predição de biodisponibilidade oral de fármacos (Kumar *et al.*, 2011). O objetivo

do algoritmo SVM é identificar um hiperplano ótimo que consiste em uma superfície de decisão linear multidimensional definida pela equação

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (3.18)$$

em que $\mathbf{w} \in \mathbb{R}^n$ é o vetor que contém os pesos (também interpretado como o vetor normal ao hiperplano), $\mathbf{x} \in \mathbb{R}^n$ é o vetor com os dados de entrada, e $b \in \mathbb{R}$ é o termo de viés (bias), responsável por deslocar o hiperplano da origem. Considerando que o espaço é ortonormal, para encontrar a distância do ponto $\mathbf{x}$ para o hiperplano, escreve-se a função

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b. \quad (3.19)$$

Utilizando a equação da projeção para representar $\mathbf{x}$ tem-se

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (3.20)$$

onde $\mathbf{x}_p$ é o vetor da projeção normal de $\mathbf{x}$ sobre o hiperplano ótimo. e r a distância entre o hiperplano ótimo e o ponto $\mathbf{x}$. Reescrevendo a equação 3.19

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{w}^T \left(\mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + b \\ &= \underbrace{\mathbf{w}^T \mathbf{x}_p}_{=-b} + r \frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|} + b = r \|\mathbf{w}\| \end{aligned} \quad (3.21)$$

logo a distância, r é

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}. \quad (3.22)$$

Para encontrar o hiperplano ótimo é necessário que o algoritmo SVM encontre os parâmetros $\mathbf{w}$ e b . Para tanto, considera-se um conjunto de treinamento $F = \{\mathbf{x}_i, d_i\}$, cujos dados sejam separáveis, onde d_i é um parâmetro binário que define em qual margem, $\Pi = \{\Pi^+, \Pi^-\}$, o vetor com os dados de entrada $\mathbf{x}_i$, será encontrado. Se d_i for igual a $+1$, $\mathbf{x}_i$ estará no lado positivo da margem (Π^+), caso d_i seja igual a -1 o vetor $\mathbf{x}_i$ estará no lado negativo da margem (Π^-).

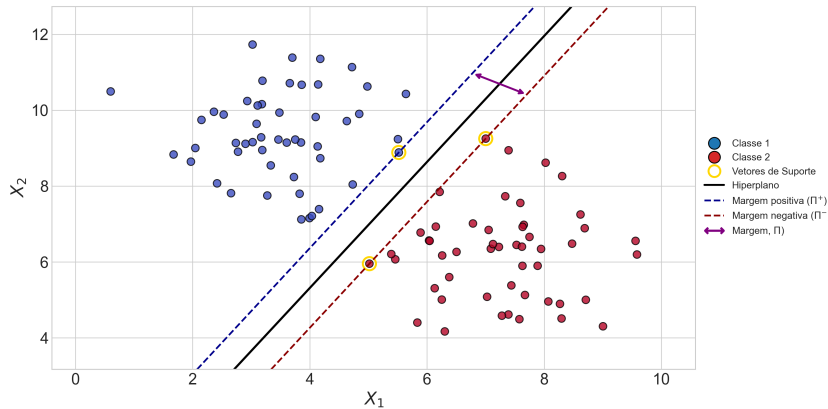
$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = \begin{cases} \geq 1, & \text{se } d_i = +1 \rightarrow \Pi^+ \\ \leq -1, & \text{se } d_i = -1 \rightarrow \Pi^- \end{cases} \quad (3.23)$$

sendo os vetores de suporte $\mathbf{x}^s$, os pontos em que a equação 3.23 tem valor igual a $+1$ ou -1

$$g(\mathbf{x}^s) = \mathbf{w}^T \mathbf{x}^s + b = \pm 1. \quad (3.24)$$

As condições matemáticas expressas nas Equações 3.23 e 3.24 encontram-se representadas geometricamente na Figura 3.2, que ilustra um cenário hipotético de classificação binária utilizando o algoritmo SVM. A visualização considera um conjunto de dados linearmente separável, composto por 100 amostras.

Figura 3.2: Resultado da simulação do algoritmo SVM para classificação binária aplicado a um conjunto de dados sintéticos linearmente separáveis. O modelo foi treinado com 100 amostras bidimensionais, com valores normalizados no intervalo $[0, 10]$.



Para equação 3.24, a distância r poderá assumir dois valores

$$r = \begin{cases} +\frac{1}{\|\mathbf{w}\|}, & \text{se } g(\mathbf{x}^s) = +1 \\ -\frac{1}{\|\mathbf{w}\|}, & \text{se } g(\mathbf{x}^s) = -1 \end{cases} \quad (3.25)$$

portanto o valor ótimo da margem de separação entre duas classes contidas no conjunto de treinamento F pode ser calculado pela diferença entre as margens

$$\Delta(\Pi) = \frac{1}{\|\mathbf{w}\|} - \left(-\frac{1}{\|\mathbf{w}\|} \right) = \boxed{\frac{2}{\|\mathbf{w}\|}}. \quad (3.26)$$

A equação 3.26 demonstra que para maximizar a margem entre duas classes no conjunto F é necessário minimizar o peso $\mathbf{w}$. No entanto, a minimização direta de $\|\mathbf{w}\|$, ou a maximização de $\frac{2}{\|\mathbf{w}\|}$ leva a um problema de otimização não diferenciável, o que inviabiliza o uso de métodos de programação quadrática ou algoritmos baseados em gradiente aplicáveis ao SVM. Para contornar essa limitação, o método primal do SVM é formulado com uma função de custo proporcional a $\|\mathbf{w}\|^2$, garantindo diferenciabilidade e permitindo a aplicação de técnicas eficientes de otimização convexa.

Método Primal

O método primal do SVM é formulado como um problema de programação quadrática convexa. Seu desenvolvimento baseia-se na minimização da norma dos pesos $\mathbf{w}$, sujeito a restrições de separabilidade das classes, caracterizando uma otimização convexa com restrições lineares. O primeiro passo do método primal consiste na reformulação da equação 3.19 com o parâmetro binário d_i , que representa a classe (rótulo) associada a cada amostra

$$g(\mathbf{x}) = d_i(\mathbf{w}^T \mathbf{x} + b), \quad d_i = \pm 1. \quad (3.27)$$

E em seguida, define-se o problema de otimização restrito a ser resolvido da seguinte forma: Dada a amostra de treinamento $\mathbf{x}_i, d_{i=1}^N$, o algoritmo deve ser capaz de encontrar os valores ótimos do vetor peso $\mathbf{w}$ e do viés b que satisfaçam a seguinte restrição

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \forall i = 1, 2, \dots, N. \quad (3.28)$$

Esta restrição impõe que todos os pontos $\mathbf{x}_i$ no conjunto de dados estejam classificados corretamente e fora da margem. Os pontos que satisfazem a restrição com igualdade — isto é, $d_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$ são os vetores de suporte $\mathbf{x}^s$, pois estão exatamente sobre as fronteiras da margem funcional. A validade dessa restrição pode ser verificada por meio da Equação 3.25, visto que, quando um dado é corretamente classificado, a restrição da Equação 3.28 será sempre satisfeita com valor positivo.

Para obter esta minimização para este problema de otimização, defini-se a função de custo para o método primal em função dos pesos $\mathbf{w}$

$$J(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}. \quad (3.29)$$

A a partir desta função de custo, constrói-se a função Lagrangiana em termos do peso, $\mathbf{w}$, do bias, b e dos multiplicadores de Lagrange α

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_i^N \alpha_i [d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (3.30)$$

onde o segundo termo do lado direito da equação, dentro dos colchetes, é obtido diretamente da restrição 3.28, ou seja, para cada restrição de desigualdade $d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 \geq 0$, é introduzido um multiplicador de Lagrange não-negativo $\alpha_i \geq 0$. Derivando a função lagrangiana em relação ao peso e ao bias e igualando os resultados a zero tem-se

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = \mathbf{w} - \sum_i^N \alpha_i d_i \mathbf{x}_i = 0 \quad (3.31)$$

$$\frac{\partial \mathcal{L}(\mathbf{w}, b, \alpha)}{\partial b} = \sum_i^N \alpha_i d_i = 0. \quad (3.32)$$

Ressalta-se que, embora a função Lagrangiana apresentada dependa explicitamente dos multiplicadores de Lagrange α_i , esses não são tratados como variáveis a serem otimizadas no problema primal, mas sim como parâmetros auxiliares introduzidos para imposição das restrições de margem. Por essa razão, não se realiza a derivação de $\mathcal{L}$ em relação a α_i (Haykin, 2001). Portanto, o método primal do algoritmo SVM (Mitchell; Mitchell, 1997), busca minimizar a função Lagrangiana, estabelecendo duas restrições:

$$\mathbf{w} = \sum_i^N \alpha_i d_i \mathbf{x}_i \quad (3.33)$$

$$\sum_i^N \alpha_i d_i = 0 \quad (3.34)$$

Método Dual

O método dual consiste em obter a minimização da função de custo, a partir exclusivamente dos multiplicadores de Lagrange, α (Haykin, 2001). Para tanto, primeiro desenvolve-se a equação 3.30

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_i^N \alpha_i d_i \mathbf{w}^T \mathbf{x}_i - b \sum_i^N \alpha_i d_i + \sum_i^N \alpha_i. \quad (3.35)$$

De acordo com as restrições estabelecidas pelo método primal, o terceiro termo da equação 3.35 é igual a zero e o primeiro termo da equação pode ser reescrito aplicando a primeira restrição do método primal em relação ao peso, $\mathbf{w}$, para função Lagrangiana

$$\|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w} = \sum_i^N \alpha_i d_i \mathbf{w}^T \mathbf{x}_i = \sum_i^N \sum_j^N \alpha_i \alpha_j d_i d_j \mathbf{x}_j^T \mathbf{x}_i \quad (3.36)$$

Aplicando a expressão 3.36, a equação 3.35 pode ser reescrita como

$$Q(\alpha) = \sum_i^N \alpha_i - \sum_i^N \sum_j^N \alpha_i \alpha_j d_i d_j \mathbf{x}_j^T \mathbf{x}_i, \quad \mathcal{L}(\mathbf{w}, b, \alpha) \rightarrow Q(\alpha) \quad (3.37)$$

portanto, no método dual o algoritmo busca apenas maximizar os valores dos multiplicadores de Lagrange. Outra vantagem da função $Q(\alpha)$ é que a sua maximização depende apenas do conjunto de dados de treinamento, $\{\mathbf{x}_i, \mathbf{x}_j^T\}_{i,j}^N$. Portanto o método dual busca maximizar os multiplicadores de Lagrange, sujeito as restrições

$$\sum_i^N \alpha_i d_i = 0 \quad (3.38)$$

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, N \quad (3.39)$$

Separação para conjunto de dados não lineares

Em caso de conjunto de dados que não sejam linearmente separáveis, adiciona-se a equação da superfície de decisão um conjunto de variáveis escalares não negativas, $\{\xi_i\}_{i=1}^N$, denominadas de variáveis soltas

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (3.40)$$

dessa forma, define-se a função de custo com um funcional de ξ

$$J(\xi) = \sum_i^N I(\xi_i - 1), \quad I(\xi) = \begin{cases} 0 & \text{se } \xi \leq 0 \\ 1 & \text{se } \xi > 1 \end{cases} \quad (3.41)$$

esta função de custo pode ser aproximada com um somatório das variáveis soltas ξ_i afim de facilitar o tratamento computacional e matemático

$$J(\xi) = \sum_{i=1}^N \xi_i. \quad (3.42)$$

A função de custo também deve ser dependente dos pesos, $\mathbf{w}$, portanto a função de custo para conjunto de dados não linearmente separáveis é formalmente definida como

$$J(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i. \quad (3.43)$$

A função $Q(\alpha)$, para o método dual é idêntica a equação para o conjunto de dados linearmente separáveis, mudando apenas o intervalo assumido pelos multiplicadores de lagrange na segunda restrição, para $0 \leq \alpha_i \leq C$, tornando os valores assumidos pelos multiplicadores de lagrange mais restritos e por consequência aprimorando os critérios de separação para um conjunto de dados não lineares.

Introdução aos Kernels em SVMs

Considera-se $\phi_j(\mathbf{x})$ como um conjunto de funções não-lineares que transferem os dados contidos no espaço de entrada para o espaço de características com dimensão infinita. A equação do hiperplano ótimo pode ser reescrita da seguinte forma

$$g(\mathbf{x}) = \sum_{j=1}^{\infty} w_j \phi_j(\mathbf{x}) = 0 \quad (3.44)$$

onde o conjunto $\{w_j\}_{j=1}^{\infty}$ representa o grupo de pesos que transformam o espaço de propriedades no espaço de saída do algoritmo. A equação 3.44 pode ser reescrita em notação matricial

$$\mathbf{w}^T \Phi(\mathbf{x}) = 0 \quad (3.45)$$

onde $\mathbf{w}$ é o vetor peso e o $\Phi(\mathbf{x})$ o vetor propriedade. Reescrevendo o vetor peso como

$$\mathbf{w} = \sum_{i=1}^{N_s} \alpha_i d_i \Phi(\mathbf{x}_i), \quad \Phi(\mathbf{x}_i) = [\phi_1(\mathbf{x}_i), \phi_2(\mathbf{x}_i), \dots]^T \quad (3.46)$$

sendo N_s o número de vetores de suporte. Combinando as expressões 3.45 e 3.46, a superfície de decisão pode ser escrita como

$$\sum_{i=1}^{N_s} \alpha_i d_i \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}) = 0. \quad (3.47)$$

O produto interno entre os vetores propriedade é formalmente definido como kernel, $\mathbf{K}$

$$\begin{aligned} \mathbf{K}(\mathbf{x}, \mathbf{x}_i) &= \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}) \\ &= \sum_{j=1}^{\infty} \phi_j(\mathbf{x}_i) \phi_j(\mathbf{x}), \quad i = 1, 2, \dots, N_s \end{aligned} \quad (3.48)$$

portanto a superfície de decisão pode ser reescrita como

$$\sum_{i=1}^{N_s} \alpha_i d_i \mathbf{K}(\mathbf{x}, \mathbf{x}_i) = 0. \quad (3.49)$$

A utilização do kernel é comumente denominada de “kernel trick”, pois especificando-o, não se faz necessário utilizar os pesos $\mathbf{w}$ para o problema de classificação. O kernel definido possui duas importantes propriedades: É uma função simétrica em relação a $\mathbf{x}_i$, $\mathbf{K}(\mathbf{x}, \mathbf{x}_i) = \mathbf{K}(\mathbf{x}_i, \mathbf{x})$, e o valor total sob a superfície da função kernel é constante. A partir desses duas propriedades, e satisfazendo o teorema de Mercer (Mercer, 1909), pode-se construir o kernel com certa flexibilidade, permitindo com que um conjunto de dados locais de menor dimensão seja não linear mas quando transportados para o espaço de característica de maior dimensão, onde ocorre a classificação, tenham comportamento linear. Abaixo são exemplificadas duas equações (polinomial e base-radial respectivamente) que podem ser aplicadas para o kernel para uma máquina de vetores de suporte

$$(\mathbf{x}^T \mathbf{x}_i + 1)^p \quad (3.50)$$

$$\exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_i\|^2\right) \quad (3.51)$$

onde a variável p na primeira expressão é a potência para o polinômio, definida pelo usuário, ou pelas condições pré-definidas do algoritmo, e a variável $\sigma^2 a$ é a largura das circunferências radiais para as funções de base radial. A construção de kernels é fundamental não apenas para o algoritmo SVM mas em toda a área de otimização em métodos de machine learning (Kung, 2014).

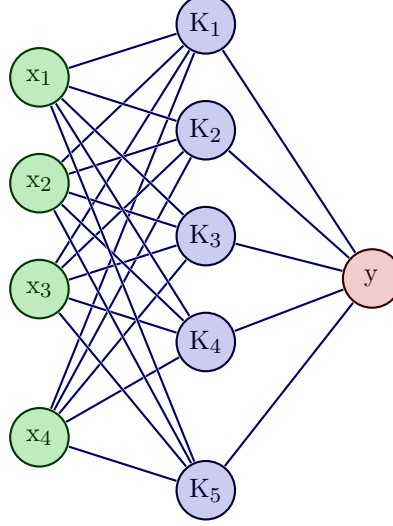


Figura 3.3: Representação de uma máquina de vetor de suporte. A primeira camada em verde é composta pelos dados de entrada contidos no vetor de entrada $\mathbf{x}$, a camada intermediária em roxo é composta por kernels com $N_s = 5$ e $\mathbf{K}_i = \mathbf{K}(\mathbf{x}, \mathbf{x}_i)$. O único neurônio de saída em vermelho corresponde a classificação (output = y) do modelo.

3.7.2 Análise discriminante gaussiana-GDA

A análise discriminante gaussiana (do inglês para *gaussian discriminant analysis*, GDA) é um algoritmo de aprendizado de máquina supervisionado, que faz parte da classe dos algoritmos de aprendizado generativo (Ng, 2008), pois efetua as predições a partir das probabilidades $P(y|X)$, sendo X o vetor aleatório que assume os valores dos dados de entrada do modelo e y a predição do modelo, diferentemente dos algoritmos de aprendizado discriminativo, como o próprio SVM, que efetuam as predições através da probabilidade $P(X|y)$. Alguns exemplos de aplicação do algoritmo GDA na área da saúde são na classificação de tumores (Gupta *et al.*, 2023), no auxílio ao diagnóstico precoce de doenças neurodegenerativas, como a DA, e em outras áreas associadas a classificação (Fang *et al.*, 2020a, 2017b,a). Os algoritmos generativos utilizam o teorema de Bayes para predição

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}. \quad (3.52)$$

O algoritmo GDA é utilizado para problemas de classificações binária, tendo suas probabilidades $P(X|y)$ e $P(y)$ distribuições normal e de Bernoulli, respectivamente

$$\begin{aligned} P(X|y) &\sim \mathcal{N}(\mu, \mathbf{C}) \\ P(y) &\sim \text{Bernoulli}(\phi) \end{aligned} \quad (3.53)$$

onde μ é o vetor média, $\mathbf{C}$ a matriz de covariância e ϕ é a probabilidade de determinada classe ser predita. A função de verossimilhança L (do inglês *likelihood function*) do algoritmo GDA para um problema de classificação binária pode ser escrita como

$$\begin{aligned} L(\phi, \mu_0, \mu_1, \mathbf{C}) &= \prod_{i=1}^m P(\mathbf{x}^i|y^i)P(y^i) \\ &= \prod_{i=1}^m \underbrace{\phi^{y^i}(1-\phi)^{(1-y^i)}}_{=P(y^i)} \underbrace{\frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}^i - \mu_k)^T \mathbf{C}^{-1}(\mathbf{x}^i - \mu_k)\right)}_{=P(\mathbf{x}^i|y^i)}, \quad k \in [0, 1] \end{aligned} \quad (3.54)$$

os parâmetros ϕ , μ_k e $\mathbf{C}$ que maximizam a função de verossimilhança L , são obtidos derivando a versão logaritmica da função L , em função de cada um dos parâmetros e igualando o resultado a zero. Os resultados obtidos são

$$\phi = \frac{\sum_{i=1}^m 1\{y^i = 1\}}{m} \quad (3.55)$$

$$\mu_k = \frac{\sum_{i=1}^m \mathbf{x}^{(i)} 1\{y^{(i)} = k\}}{\sum_{i=1}^m 1\{y^{(i)} = k\}}, \quad k \in [0.1] \quad (3.56)$$

$$\mathbf{C} = \frac{1}{m} \sum_{i=1}^m (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)^T \quad (3.57)$$

onde o termo $1\{y^i = k\}$ corresponde a função indicadora.

3.8 Algoritmos de clusterização

3.8.1 Kmeans

O algoritmo kmeans é um método de aprendizado de máquina não supervisionado, fazendo parte dos métodos de clusterização. Foi desenvolvido em 1957 por Stuart Lloyd quando trabalhava na Bell Labs (Lloyd, 1957). O algoritmo de Lloyd é a implementação computacional mais empregada, todavia existem outras implementações, como a de MacQueen que cunhou o termo kmeans para este algoritmo (MacQueen *et al.*, 1967). O kmeans é um algoritmo amplamente utilizado no drug design para a clusterização dos fármacos e para análise de dados associados a estes (Lu *et al.*, 2016; Hananto *et al.*, 2021; Madugula *et al.*, 2021). O objetivo do algoritmo é determinar o conjunto de centroides $\mathbf{C}$ para cada cluster, sendo $\mathbf{A}$ um conjunto de dados no espaço $\mathbb{R}^d$. Para aplicar o algoritmo, o conjunto de dados $\mathbf{A}$ é decomposto em k subconjuntos disjuntos (clusters)

$$\mathbf{A} = \mathbf{A}^1 \cup \mathbf{A}^2 \cup \dots \cup \mathbf{A}^k \quad (3.58)$$

sendo o conjunto de centroides $\mathbf{C}$

$$\mathbf{C} = \langle \mathbf{c}^1, \mathbf{c}^2, \dots, \mathbf{c}^k \rangle \quad (3.59)$$

que minimizam a função de custo, $J(\mathbf{A}, \mathbf{C})$, que para o algoritmo kmeans é definida como o somatório das distâncias entre cada vetor propriedade, $\mathbf{x}$, com o centróide mais próximo

$$J(\mathbf{A}, \mathbf{C}) = \sum_{a=1}^k \sum_{\mathbf{x} \in \mathbf{A}^i} \Delta(\mathbf{x}, \mathbf{c}^i). \quad (3.60)$$

O teorema central do algoritmo determina que a média de determinado cluster $\mathbf{A}^i$ é o único valor possível para o centróide $\mathbf{c}^i$ que minimiza a função de custo (Bishop; Nasrabadi, 2006). É possível provar este teorema desenvolvendo a função de custo J para o centróide em questão

$$\begin{aligned} J(\mathbf{A}^i, \mathbf{C}) &= \sum_{\mathbf{x} \in \mathbf{A}^i} \Delta(\mathbf{x}, \mathbf{c}^i) = \sum_{\mathbf{x} \in \mathbf{A}^i} \sum_{a=1}^d (x_a - c_a^i)^2 \\ &= \sum_{a=1}^d \left(\left(\sum_{\mathbf{x} \in \mathbf{A}^i} x_a^2 \right) - 2 \left(\sum_{\mathbf{x} \in \mathbf{A}^i} x_a \right) c_a^i + \underbrace{\sum_{\mathbf{x} \in \mathbf{A}^i} (c_a^i)^2}_{=|\mathbf{A}^i|(c_a^i)^2} \right) \end{aligned}$$

e em seguida definindo a derivada em relação ao centróide c_a^i igual a zero

$$\frac{\partial}{\partial c_a^i} J(\mathbf{A}^i, \mathbf{C}) = -2 \sum_{\mathbf{x} \in \mathbf{A}^i} x_a + 2|\mathbf{A}^i|c_a^i = 0$$

então, para cada centróide c_a^i , têm-se que

$$c_a^i = \frac{1}{|A^i|} \sum_{x \in A^i} x_a$$

que pode ser generalizada para o vetor centróide associado ao cluster A^i

$$\mathbf{c}^i = \frac{1}{|A^i|} \sum_{x \in A^i} \mathbf{x} \quad (3.61)$$

então, defini-se o centróide $\mathbf{c}^i$ como um ponto d-dimensional onde seus valores, em cada dimensão são a média dos valores de todos os pontos no cluster A^i associado.

3.9 Visão Computacional

3.9.1 Introdução

A área da visão computacional pode ser definida como o campo de estudo que tem por objetivo extrair, processar, analisar dados de imagens digitais e produzir, a partir da análise destes dados, informação para um determinado objetivo (Klette, 2014). Possui um extenso campo de aplicações como na automação de veículos (Janai *et al.*, 2020), no processamento de imagens médicas para diagnóstico (Esteve *et al.*, 2021), no reconhecimento facial em sistemas de segurança (Karnila; Irianto, 2024), na detecção de objetos específicos em imagens (Cazzato *et al.*, 2020), no controle de qualidade de empresas (Islam *et al.*, 2024), dentre outras diversas aplicações.

3.9.2 Métodos utilizados: GLCM

O método GLCM (do inglês *gray-level-co-occurrence-matrix*) é uma importante ferramenta para extração da informação de imagens, representando a informação estatística de segunda ordem dos níveis de cinza entre a vizinhança de pixels em uma imagem (Zubair; Alo, 2024; Johnson; Baker Jr, 2004; Tadmor; Tolhurst, 1994). A partir da matriz GLCM, é possível obter descrições estatísticas, como energia, entropia, contraste e homogeneidade, denominadas de texturas da imagem (Elvitaria *et al.*, 2024; Lubis; Siregar *et al.*, 2024). O algoritmo que gera a matriz GLCM, introduzido inicialmente por Haralick e colaboradores (Haralick; Shanmugam, s.d.), funciona da seguinte forma: dada uma imagem em escala de cinza, representada como uma matriz cujos elementos são os pixels da imagem, $I(x, y)$, onde x e y são as coordenadas espaciais, o operador de co-ocorrência $C(i, j)$, computa a frequência com que pares de pixels com valores específicos i e j , separados por um deslocamento espacial pré-determinado $(\Delta x, \Delta y)$, ocorrem na imagem. A Equação 3.62 é uma versão adaptada, da abordagem descrita, desenvolvida posteriormente a Haralick, por trabalhos subsequentes, como o de Gonzalez e Woods (Gonzalez, 2009).

$$C(i, j) = \sum_{x=1}^n \sum_{y=1}^m \begin{cases} 1, & \text{se } I(x, y) = i \wedge I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{caso contrário} \end{cases} \quad (3.62)$$

A Equação 3.63 demonstra o processo de construção da matriz GLCM. Dada uma imagem I , o operador C realiza uma varredura sistemática para contabilizar a frequência de ocorrência dos pares de pixels na imagem. A matriz central da expressão exemplifica este procedimento de mapeamento, considerando $\Delta x = 1$ e $\Delta y = 0$ (deslocamento horizontal à direita, com pares de pixels adjacentes a uma distância de 1 pixel)

$$C(I) = C \begin{pmatrix} 1 & 0 & 0 & 2 \\ 2 & 2 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \boxed{0} & \boxed{0} & \boxed{1} & \boxed{2} \\ \boxed{0} & (0,0) & (0,1) & (0,2) \\ \boxed{1} & (1,0) & (1,1) & (1,2) \\ \boxed{2} & (2,0) & (2,1) & (2,2) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 3 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix}}_{\text{matriz GLCM}} \quad (3.63)$$

para ilustrar o processo, considere a notação $(0,0)$: esta representa a posição na matriz de co-ocorrência onde será armazenado o número de vezes que pixels com valores 0 e 0 aparecem adjacentes (considerando o deslocamento especificado) em toda a imagem I . No exemplo apresentado, esta configuração ocorre apenas uma vez, especificamente na primeira linha da imagem original. Consequentemente, a posição $(0,0)$ na

matriz GLCM resultante recebe o valor 1, refletindo esta única ocorrência do par de pixels. Portanto, a matriz GLCM constitui uma ferramenta eficaz para a caracterização de propriedades texturais da imagem, permitindo a quantificação de padrões de co-ocorrência e transições de intensidade que definem a estrutura espacial dos pixels.

Tabela 3.1: Principais propriedades estatísticas de textura de imagens

Função	Expressão
Contraste	$\sum_{i=1}^N \sum_{j=1}^N (i-j)^2 P(i,j)$
Entropia	$-\sum_{i=1}^N \sum_{j=1}^N P(i,j) \log P(i,j)$
Correlação	$\frac{\sum_{i=1}^N \sum_{j=1}^N (i-\mu_x)(j-\mu_y) P(i,j)}{\sigma_x \sigma_y}$
Energia	$\sum_{i=1}^N \sum_{j=1}^N P(i,j)^2$

3.9.3 Métodos utilizados: Filtro de Sobel

O filtro de Sobel é utilizado em visão computacional para a detecção de arestas, ou contornos em imagens (Dharampal, 2015). O operador ou filtro de Sobel (S), é uma operação de convolução (*), em uma imagem **A** utilizando matrizes de dimensão 3x3 (**G_y**, **G_x**). Essa operação gera espécies de “derivadas” da imagem, sendo uma na vertical (eixo y) e outra na horizontal (eixo x)

$$S_y(\mathbf{A}) = \mathbf{G}_y \mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} * \mathbf{A} \quad (3.64)$$

$$S_x(\mathbf{A}) = \mathbf{G}_x \mathbf{A} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * \mathbf{A} \quad (3.65)$$

sendo o tamanho do vetor **G** e a direção **Θ** definidos como

$$\mathbf{G} = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2} \quad (3.66)$$

$$\Theta = \arctan\left(\frac{\mathbf{G}_y}{\mathbf{G}_x}\right) \quad (3.67)$$

3.10 Métodos para tratamento dos dados

3.10.1 Análise dos componentes principais-PCA

A análise dos componentes principais (PCA) foi formulada em 1901 pelo estatístico Karl Pearson (Pearson, 1901), e consiste na diminuição da dimensionalidade de um conjunto de dados, com perda mínima da informação contida nestes. A PCA é uma ferramenta importante em ciência de dados, sendo aplicada para o tratamento e na identificação das variáveis (descritores no caso dos fármacos) de maior relevância em determinado conjunto de dados. Dado um conjunto de dados $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_N\}$, $\mathbf{x}_N \in \mathbb{R}$, o objetivo do algoritmo moderno da PCA, assim como os demais algoritmos de machine learning, é encontrar os valores das variáveis $\mathbf{x}_n$ que minimizem a função de custo, J (Bishop; Nasrabadi, 2006; Hastie *et al.*, 2009; Jolliffe, 2002).

$$J = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2 \quad (3.68)$$

onde N é o número total do conjunto de dados, $\mathbf{x}_n$ as variáveis ou predições corretas e $\tilde{\mathbf{x}}_n$ as variáveis aproximadas. A variável aproximada $\tilde{\mathbf{x}}_n$ pode ser definida como o somatório do produto entre as projeções

ortonormais C_{in} dos dados de entrada com os vetores de base ortogonais $\mathbf{b}_i$, em todo espaço amostral (Watkins, 2004; Hastie *et al.*, 2009)

$$\tilde{\mathbf{x}}_n = \sum_{j=1}^M C_{jn} \mathbf{b}_j + \sum_{i=M+1}^D C_{in} \mathbf{b}_i \quad (3.69)$$

na PCA o segundo termo, que é o complemento ortogonal ao primeiro termo da equação 3.69, é ignorado, tratando-se exclusivamente o primeiro termo, com os dados contidos no denominado subespaço principal M. A função de custo deve ser minimizada em relação as projeções ortonormais C_{in} no subespaço principal, uma vez que estas determinam os valores assumidos pelas variáveis aproximadas $\tilde{\mathbf{x}}_n$, enquanto os vetores de base $\mathbf{b}_i$ tem valores bem definidos de acordo com o subespaço amostral. Portanto, defini-se como condição para minimização que a derivada parcial da função de custo em relação as projeções ortonormais C_{in} seja igual a zero

$$\frac{\partial J}{\partial C_{in}} = 0 \quad (3.70)$$

usando a derivação em cadeia, têm-se

$$\frac{\partial J}{\partial C_{in}} = \frac{\partial J}{\partial \tilde{\mathbf{x}}_n} \frac{\partial \tilde{\mathbf{x}}_n}{\partial C_{in}} \quad (3.71)$$

resolvendo o primeiro termo da expressão 3.71

$$\frac{\partial J}{\partial \tilde{\mathbf{x}}_n} = \frac{1}{N} \frac{\partial}{\partial \tilde{\mathbf{x}}_n} \sum_{i=1}^N [(\mathbf{x}_n - \tilde{\mathbf{x}}_n)^T (\mathbf{x}_n - \tilde{\mathbf{x}}_n)]^2 = -\frac{2}{N} (\mathbf{x}_n - \tilde{\mathbf{x}}_n)^T \quad (3.72)$$

e para o segundo termo

$$\frac{\partial \tilde{\mathbf{x}}_n}{\partial C_{in}} = \frac{\partial}{\partial C_{in}} \sum_{j=1}^M C_{jn} \mathbf{b}_j = \mathbf{b}_i \quad (3.73)$$

utilizando as expressões obtidas em 3.72 e 3.73 na equação 3.71

$$\frac{\partial J}{\partial C_{in}} = -\frac{2}{N} (\mathbf{x}_n - \tilde{\mathbf{x}}_n)^T \mathbf{b}_i = -\frac{2}{N} (\mathbf{x}_n - \sum_{j=1}^M C_{jn} \mathbf{b}_j)^T \mathbf{b}_i = 0 \quad (3.74)$$

desenvolvendo esta expressão e aplicando a ortonormalidade dos vetores de base $\mathbf{b}_i$

$$\frac{\partial J}{\partial C_{in}} = -\frac{2}{N} \mathbf{x}_n^T \mathbf{b}_i + \frac{2}{N} \underbrace{\sum_{j=1}^M C_{jn} \mathbf{b}_j^T \mathbf{b}_i}_{=C_{in} \underbrace{\mathbf{b}_i^T \mathbf{b}_i}_{=1}} = -\frac{2}{N} \mathbf{x}_n^T \mathbf{b}_i + \frac{2}{N} C_{in} = 0. \quad (3.75)$$

Portanto, o coeficiente C_{in} , que projeta a variável $\mathbf{x}_n$ no subespaço principal e minimiza a função de custo J para PCA corresponde a equação

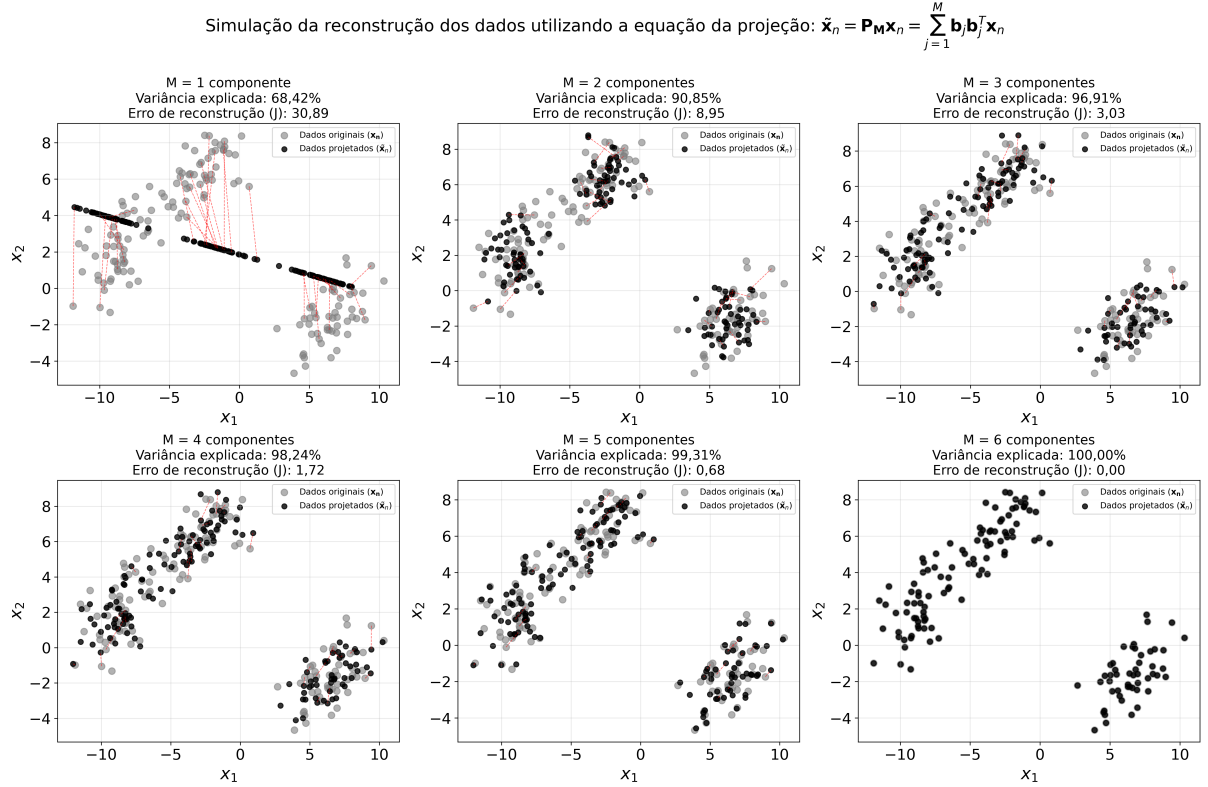
$$C_{in} = \mathbf{x}_n^T \mathbf{b}_i \quad (3.76)$$

reescrevendo $\tilde{\mathbf{x}}_n$ como

$$\tilde{\mathbf{x}}_n = \sum_{j=1}^M (\mathbf{x}_n^T \mathbf{b}_j) \mathbf{b}_j = \sum_{j=1}^M \mathbf{b}_j (\mathbf{b}_j^T \mathbf{x}_n) = \underbrace{\sum_{j=1}^M (\mathbf{b}_j \mathbf{b}_j^T)}_{=\mathbf{P}_M} \mathbf{x}_n = \mathbf{P}_M \mathbf{x}_n \quad (3.77)$$

onde $\mathbf{P}_M$ é a matriz de projeção. A Figura 3.4 ilustra o processo de reconstrução descrito pela Equação 3.77, demonstrando como diferentes valores de M afetam a qualidade da aproximação $\tilde{\mathbf{x}}_n$.

Figura 3.4: Demonstração da reconstrução de dados via PCA com diferentes números de componentes principais (M). Os pontos cinza representam os dados originais ($\mathbf{x}_n$) e os pontos pretos mostram os dados reconstruídos ($\tilde{\mathbf{x}}_n$) após projeção no subespaço de dimensão M. As linhas tracejadas vermelhas conectam cada ponto original à sua reconstrução, ilustrando o erro associado a projeção. A variância explicada e o erro de reconstrução (J) são mostrados para cada valor de M



A partir da Equação 3.77, pode-se concluir que a diferença $|\mathbf{x}_n - \tilde{\mathbf{x}}_n|$ será igual a diferença entre as matrizes de projeção

$$\underbrace{\mathbf{P}_M \mathbf{x}_n}_{=\mathbf{x}_n} + \mathbf{P}_{M \rightarrow D} \mathbf{x}_n - \underbrace{\mathbf{P}_M \mathbf{x}_n}_{=\tilde{\mathbf{x}}_n} = \mathbf{P}_{M \rightarrow D} \mathbf{x}_n \quad (3.78)$$

aplicando este resultado a função de custo:

$$J = \frac{1}{N} \sum_{n=1}^N |\mathbf{x}_n - \tilde{\mathbf{x}}_n|^2 = \frac{1}{N} \sum_{n=1}^N (\mathbf{P}_{M \rightarrow D} \mathbf{x}_n)^2 = \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D [(\mathbf{b}_i \mathbf{b}_i^T) \mathbf{x}_n]^2 \quad (3.79)$$

rearranjando a última expressão para obter a matriz de covariância $\mathbf{S}$

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D [(\mathbf{b}_i \mathbf{b}_i^T) \mathbf{x}_n]^2 &= \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D (\mathbf{b}_i^T \mathbf{x}_n)^2 \underbrace{\mathbf{b}_i^2}_{=1} \\ &= \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D (\mathbf{b}_i^T \mathbf{x}_n)^2 \\ &= \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D (\mathbf{b}_i^T \mathbf{x}_n)(\mathbf{x}_n^T \mathbf{b}_i). \end{aligned}$$

Portanto, a função de custo pode ser reescrita

$$J = \sum_{i=M+1}^D \mathbf{b}_i^T \underbrace{\left(\frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^T \right)}_{=\mathbf{S}} \mathbf{b}_i = \sum_{i=M+1}^D \mathbf{b}_i^T \mathbf{S} \mathbf{b}_i. \quad (3.80)$$

Para otimizar os vetores de base $\mathbf{b}_j$, utiliza-se a técnica dos multiplicadores de Lagrange (Bishop; Nasrabadi, 2006; Jolliffe, 2002; Trefethen; Bau, 2022). No caso de um sistema bi-dimensional, considere-se $\mathbf{b}_1$ e $\mathbf{b}_2$ vetores unitários, com $\mathbf{b}_1$ pertencendo ao subespaço principal e $\mathbf{b}_2$ ao subespaço complementar. Primeiro, a função de custo é escrita em termos de $\mathbf{b}_2$:

$$\mathbf{J} = \mathbf{b}_2^T \mathbf{S} \mathbf{b}_2, \quad \mathbf{b}_1^T \mathbf{b}_j = \begin{cases} \delta_{ij} \delta_{ij} = 1, \text{ se } i = j \\ \delta_{ij} \delta_{ij} = 0, \text{ caso contrário} \end{cases} \quad (3.81)$$

onde δ_{ij} é o delta de kronecker. Escrevendo a função lagrangiana para $\mathbf{b}_2$ e igualando a derivada parcial da função lagrangiana $\mathcal{L}(\mathbf{J})$ em termos de $\mathbf{b}_2$ a zero:

$$\begin{aligned} \mathcal{L}(\mathbf{J}) &= \mathbf{b}_2^T \mathbf{S} \mathbf{b}_2 + \lambda(1 - \mathbf{b}_2^T \mathbf{b}_2), \\ \frac{\partial \mathcal{L}(\mathbf{J})}{\partial \mathbf{b}_2} &= 2\mathbf{b}_2^T \mathbf{S} - 2\lambda \mathbf{b}_2^T = 0. \end{aligned} \quad (3.82)$$

A última equação pode ser escrita como uma equação de autovalor $\mathbf{S} \mathbf{b}_2 = \lambda \mathbf{b}_2$, podendo ser generalizada para N-dimensões

$$\mathbf{S} \mathbf{b}_j = \lambda_j \mathbf{b}_j, \quad j = M + 1, \dots, D \quad (3.83)$$

onde $\mathbf{b}_j$ é a j -ésima componente principal (autovetor) e λ_j o autovalor correspondente. Aplicando a Equação de autovalor 3.83 para função de custo

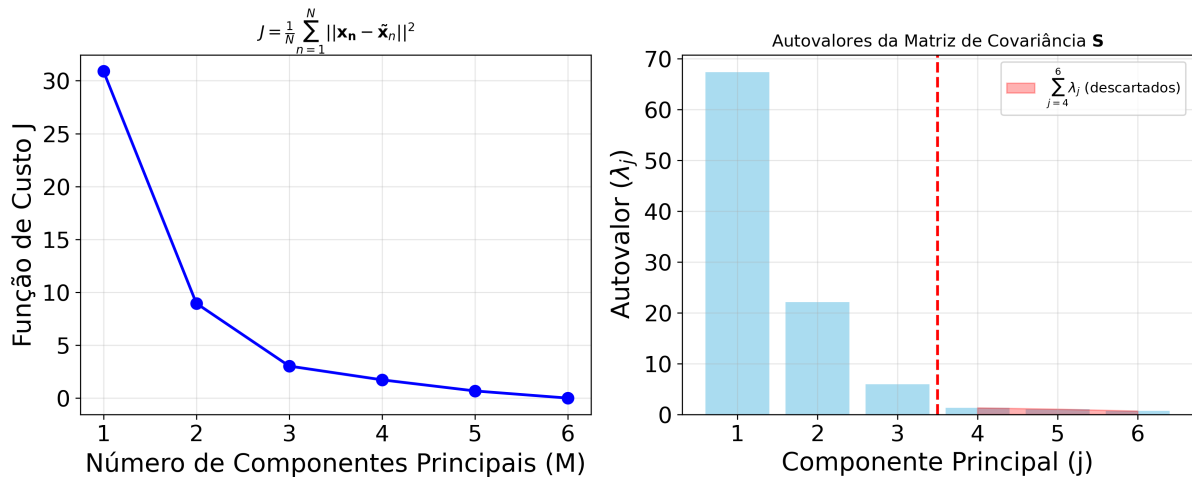
$$\begin{aligned} \mathbf{J} &= \sum_{j=M+1}^D \mathbf{b}_j^T \mathbf{S} \mathbf{b}_j = \sum_{j=M+1}^D \mathbf{b}_j^T \lambda_j \mathbf{b}_j \\ &= \sum_{j=M+1}^D \lambda_j \underbrace{\mathbf{b}_j \mathbf{b}_j^T}_{=1} \end{aligned}$$

portanto, a função de custo pode ser escrita como o somatório dos autovalores λ_j no subespaço complementar (Jolliffe, 2002; Abdi; Williams, 2010)

$$\boxed{\mathbf{J} = \sum_{j=M+1}^D \lambda_j.} \quad (3.84)$$

A Figura 3.5 , que se baseia na simulação feita na Figura 3.4, demonstra empiricamente a relação estabelecida na Equação 3.84, evidenciando que a função de custo $\mathbf{J}$ corresponde exatamente à soma dos autovalores descartados no subespaço complementar.

Figura 3.5: Esquerda: Função de custo $\mathbf{J}$ versus número de componentes principais M para simulação da Figura 3.4, mostrando a redução do erro de reconstrução conforme mais componentes são incluídos; Direita: Autovalores da matriz de covariância, onde cada autovalor λ_j representa a variância explicada pelo j -ésimo componente principal. A área vermelha destaca os autovalores descartados quando $M = 3$, que correspondem diretamente ao valor da função de custo $\mathbf{J}$.



3.11 Redes neurais artificiais

3.11.1 Introdução

O primeiro modelo que emulava uma estrutura neuronal computacionalmente foi criado pelos cientistas da área da neurociência computacional, Warren McCulloch e Walter Pitts em 1943, com o objetivo de empregá-lo na análise de padrões na informação (McCulloch; Pitts, 1943). Entretanto, outros pesquisadores foram cruciais para o desenvolvimento da área, como o cientista da computação Alan Turing, que em 1935 descreveu uma máquina de computação abstrata com memória ilimitada, contendo um leitor capaz de escanear toda a memória, símbolo à símbolo (Turing, 1935). Warren McCulloch e Walter Pitts inventaram um perceptron, que consiste em um classificador binário que mapeia os dados de entrada contidos no vetor $\mathbf{x}$, com os pesos $\mathbf{w}$ e o bias associado $\mathbf{b}$ do modelo, para um valor binário de saída por meio da função $f(\mathbf{x})$

$$f(\mathbf{x}) = \begin{cases} 1 & \text{se } \mathbf{w}^T \mathbf{x} + \mathbf{b} \geq 0 \\ 0 & \text{caso contrário.} \end{cases} \quad (3.85)$$

O perceptron é o neurônio da rede neural artificial. Atualmente diversas funções de ativação podem ser empregadas em um neurônio artificial para converterem a função $f(\mathbf{x})$ em um valor de saída y para o modelo, como softmax (Bridle, 1989), sigmoide (Verhulst, 1847), relu (Nair; Hinton, 2010) e tanh (Bradley *et al.*, 2007).

Tabela 3.2: Principais funções de ativação utilizadas em redes neurais.

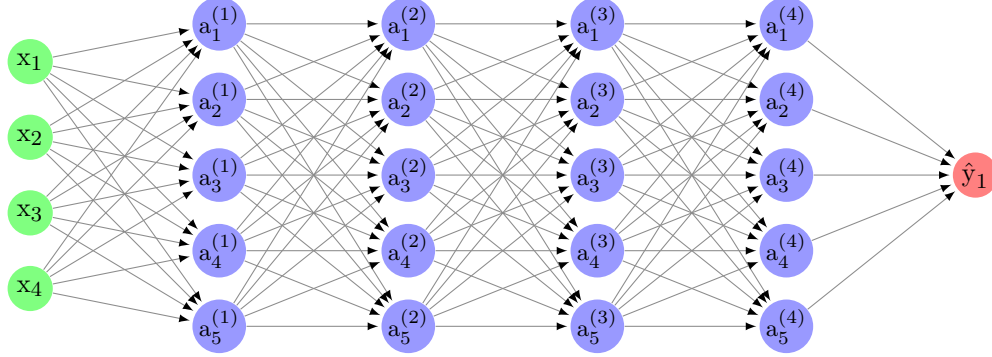
Função	Expressão	Intervalo
sigmoide	$\sigma(x) = \frac{1}{1+e^{-x}}$	$(0, 1)$
softmax	$s(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$	$(0, 1)$
ReLU	$R(x) = \max(0, x)$	$(0, \infty)$
tanh	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$(-1, 1)$

Outra importante contribuição para o campo da inteligência artificial deve-se ao neurocientista canadense Donald Hebb (Hebb, 1949), que postulou a denominada lei de Hebb em 1949, a qual afirma que, se dois neurônios estiverem ativados, a conexão entre eles é reforçada e caso apenas um esteja ativado, a conexão entre eles é enfraquecida. Apesar da lei de Hebb não ser completamente aceita como mecanismo de aprendizado em humanos, ainda é utilizada como uma das regras básicas utilizadas durante o treinamento de redes neurais artificiais.

Após uma diminuição do interesse na área no final dos anos 60, a partir de 1980, os estudos na área retornaram com grande impacto. Em 1982, o físico John Hopfield publicou seu trabalho sobre memória associativa (Hopfield, 1982). Ele propôs que as memórias não eram armazenadas de forma isolada, mas sim como padrões interconectados de atividade neural. Ele desenvolveu um modelo para simular este processo, denominado redes de Hopfield. O psicólogo cognitivo e cientista da computação Geoffrey Hinton, aperfeiçoou o modelo de Hopfield em 1985, utilizando a equação de Boltzmann para desenvolver a chamada máquina de Boltzmann (Ackley *et al.*, 1985).

A principal diferença entre os modelos, é que a arquitetura da máquina de Boltzmann é assimétrica, ou seja, cada camada da rede não possui necessariamente o mesmo número de neurônios. A máquina de Boltzmann possui neurônios visíveis que recebem os dados e neurônios ocultos (*hidden neurons*) que determinam o funcionamento das redes como um todo. Esta arquitetura permite que o modelo aprenda a partir de exemplos, enquanto as redes de Hopfield eram capazes apenas de reconhecer padrões. As máquinas de Boltzmann são a base para as redes neurais recorrentes (RNR). As contribuições de Hopfield e de Hinton lhes renderam o prêmio nobel de Física de 2024 (Nobel, 2024).

Figura 3.6: Representação de uma rede neural composta por: uma camada de entrada em verde, que recebe os dados x_i , quatro camadas ocultas com o neurônio com a função de ativação $a_i = g(\mathbf{w}\mathbf{x}_i + \mathbf{b})$, onde $g(\cdot)$ é uma das funções da tabela Tabela 3.1, e uma camada contendo um único neurônio com a saída $\hat{y}$ da rede.



3.11.2 Retropropagação

O algoritmo de retropropagação é um dos algoritmos de aprendizado de máquina mais importantes para o treinamento de redes neurais artificiais. Foi implementado pela primeira vez pelo matemático e cientista da computação finlandês Seppo Linnainmaa em sua dissertação de mestrado em 1970 sobre o desenvolvimento de métodos de forma analítica e algorítmica para determinação de coeficientes da expansão de Taylor associados aos erros de arredondamento locais (Linnainmaa, 1970). Todavia o algoritmo passou a ser popularizado para o treinamento de redes neurais artificiais com os pesquisadores Rumelhart, Hinton e Williams, com seu artigo de 1986 sobre o aprendizado de representações através da retropropagação do erro associado ao valor de saída de uma rede neural artificial (Rumelhart *et al.*, 1986). O objetivo do algoritmo de retropropagação é minimizar o erro associado ao valor de saída de modelo, através da atualização dos pesos ($\mathbf{w}$) e bias ($\mathbf{b}$). Para que a rede efetue a atualização de seus pesos e bias em cada iteração, ela deve retornar às camadas ocultas, a partir da camada de saída. Este processo é realizado através dos gradientes da função de custo J do modelo em relação aos pesos, funções de ativação e bias. Para o gradiente em relação aos pesos $w_{jk}^{(C)}$ tem-se

$$\nabla J(w_{jk}^{(C)}) = \frac{\partial J}{\partial a_j^{(C)}} \frac{\partial a_j^{(C)}}{\partial z_j^{(C)}} \frac{\partial z_j^{(C)}}{\partial w_{jk}^{(C)}} = \frac{\partial J}{\partial w_{jk}^{(C)}} \quad (3.86)$$

onde C é o número da camada, j e k são os índices dos neurônios das camadas $C-1$ e C , respectivamente, $z_j^{(C)}$ a função de pré-ativação e w_{jk} o peso associado aos neurônios j e k . A função de pré-ativação $z_j^{(C)}$ efetua um somatório sobre os produtos formados entre os pesos $w_{jk}^{(C)}$ com os dados de entrada da camada anterior $x^{(C-1)}$

$$z_j^{(C)} = \sum_{i=0}^n w_{ji} x^{(C-1)} \quad (3.87)$$

onde n é o número de neurônios na camada $C-1$. O cálculo do gradiente nas camadas ocultas envolve a derivação parcial da função de custo em relação a cada função de ativação de forma similar as expressões anteriores. No entanto, a função de ativação $a_k^{(C-1)}$ de um neurônio afeta os resultados dos demais neurônios da próxima camada aos quais se conecta, adicionando-se um somatório a expressão de seu gradiente

$$\nabla J(a_k^{(C-1)}) = \sum_{j=0}^n \frac{\partial J}{\partial a_j^{(C)}} \frac{\partial a_j^{(C)}}{\partial z_j^{(C)}} \frac{\partial z_j^{(C)}}{\partial a_k^{(C-1)}} = \frac{\partial J}{\partial a_k^{(C-1)}}. \quad (3.88)$$

O gradiente da função de custo em relação ao bias $b_j^{(C)}$ é calculado da mesma forma que em relação aos pesos, sendo atualizado a cada iteração

$$\nabla J(b_j^{(C)}) = \frac{\partial J}{\partial a_j^{(C)}} \frac{\partial a_j^{(C)}}{\partial z_j^{(C)}} \frac{\partial z_j^{(C)}}{\partial b_j^{(C)}} = \frac{\partial J}{\partial b_j^{(C)}} \quad (3.89)$$

3.11.3 Autoencodificador variacional- VAE

O Modelo autoencodificador variacional (do inglês *variational autoencoder*, VAE) é uma técnica avançada de aprendizado de máquina profundo, amplamente utilizada para geração de dados, redução de dimensionalidade e modelagem de distribuições complexas. Sua aplicação na quimioinformática é particularmente relevante, pois permite a geração eficiente de novas moléculas, agilizando o processo de descoberta de substâncias com potencial terapêutico (Gircha *et al.*, 2021; Mukesh *et al.*, 2022; Bhadwal *et al.*, 2024; Hu *et al.*, 2023a; Huang *et al.*, 2023). A arquitetura de um modelo VAE é composto por dois componentes interconectados: um encodificador que recebe os dados de entrada x , gerando uma distribuição a posteriori $q_\phi(z|x)$ no espaço latente, representado pelo vetor z , e um decodificador que utiliza a distribuição de verossimilhança (*likelihood*), $p_\theta(x|z)$, para gerar a saída do modelo. O objetivo do modelo VAE é minimizar a perda entre os dados originais e os dados reconstruídos pelo modelo. A função de custo de um modelo VAE é definida como

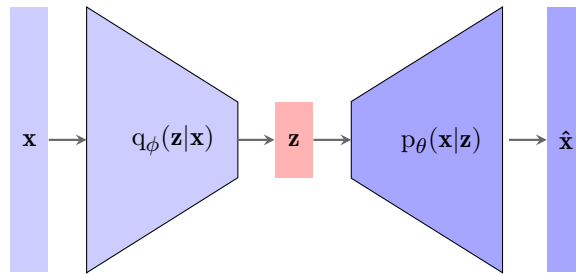
$$\mathcal{L}(x^{(i)}, \theta, \phi) = \mathbb{E}_z \left[\log p_\theta(x^{(i)}|z) \right] - D_{KL} \left(q_\phi(z|x^{(i)}) \parallel p_\theta(z) \right) \quad (3.90)$$

onde θ indica os parâmetros do decodificador e ϕ os parâmetros do encodificador. O termo $p_\theta(z)$ corresponde a probabilidade a priori, que geralmente segue uma distribuição gaussiana $p_\theta(z) \sim \mathcal{N}(0, I)$, sendo I a matriz identidade, e D_{KL} é a divergência de Kullback-Leibler, definida como

$$D_{KL}(q||p) = \mathbb{E}_{X \sim q} \left[\log \frac{q(x)}{p(x)} \right] = \mathbb{E}_{X \sim q} [\log q(x) - \log p(x)] \quad (3.91)$$

onde X é uma variável aleatória. Os modelos VAE são usados em outras áreas relacionadas a saúde como na geração de séries temporais de dados médicos (Li *et al.*, 2023), modelagem generativa de determinantes de dosagem de medicamentos para doenças renais, hepáticas, metabólicas e cardíacas (Titar; Ramanathan, 2024) e na detecção de patologias em imagens médicas (Uzunova *et al.*, 2019).

Figura 3.7: Representação de um autoencodificador com entrada x , processamento e saída reconstruída $\hat{x}$



3.11.4 Redes neurais convolucionais

Redes neurais convolucionais (RNCs) são uma arquitetura de rede neural profunda projetada para o processamento de dados multidimensionais, especialmente imagens (Li *et al.*, 2021). Possuem basicamente três tipos de camadas: convolucionais, de agrupamento (*pooling*) e densas. As camadas convolucionais aplicam operações de convolução, utilizando filtros para detecção de padrões locais (Liu *et al.*, 2015). Já as camadas de agrupamento efetuam operações de redução das dimensões espaciais do mapa de características, reduzindo a dimensionalidade e ampliando o campo receptivo da imagem (Zhang *et al.*, 2020; Gao *et al.*, 2019). As camadas densas seguem a tradicional arquitetura de RNA. As RNCs utilizam funções de ativação, como ReLU, Sigmoid e Tanh, e são treinadas por meio de retropropagação e otimização de parâmetros através de algoritmos de otimização, como SGD e Adam. O operador de convolução (*) é uma operação empregada em matemática, nas áreas de análise funcional e processamento

de sinais, que gera uma função $((f * g)(x))$ a partir de duas funções reais $(f(x), g(x))$, medindo a soma do produto entre as duas funções originais ao longo da região de superposição entre elas.

$$(f * g)(x) := \int_{-\infty}^{\infty} f(u)g(x - u)du \quad (3.92)$$

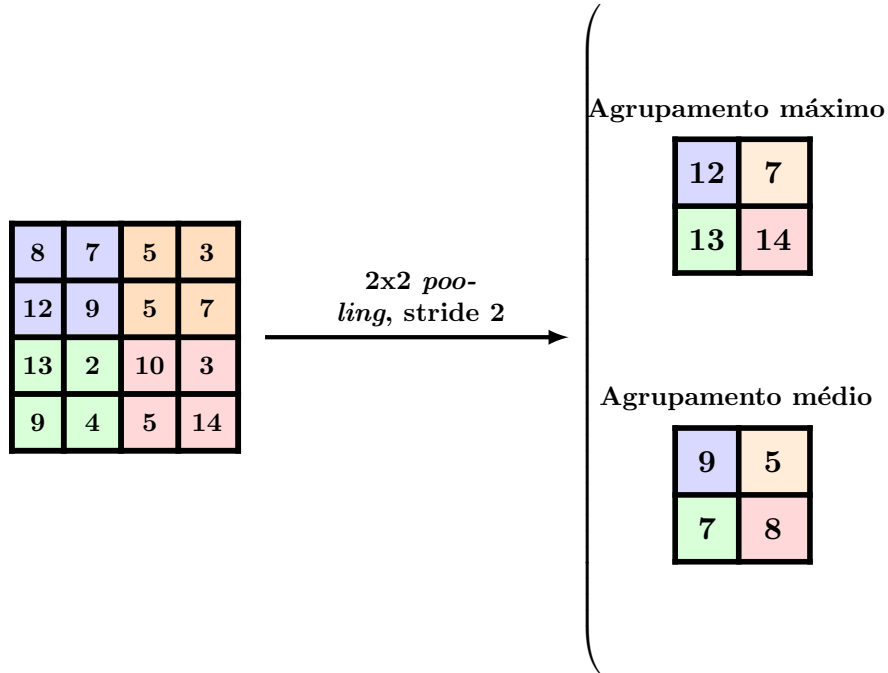
onde x é a variável independente (geralmente representada pelo tempo), e u o deslocamento.

Figura 3.8: Representação de uma operação de convolução

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 & 4 & 1 \\ 1 & 2 & 4 & 3 & 3 \\ 1 & 2 & 3 & 4 & 1 \\ 1 & 3 & 3 & 1 & 1 \\ 3 & 3 & 1 & 1 & 0 \end{pmatrix}$$

I K I * K

Figura 3.9: Representação de dois tipos de operações de *pooling* para redução de dimensionalidade do mapa de características. O hiperparâmetro *stride* define a distância percorrida pelo filtro, influenciando a redução dimensional.



Dentre as aplicações das RNCs na área de saúde incluem-se, o diagnóstico de doenças (Yadav; Jadhav, 2019), análise de imagens radiológicas (Soffer *et al.*, 2019), detecção e estágios de câncer de pulmão (Chon *et al.*, 2017), mama (Ting *et al.*, 2019) e tumores cerebrais (Abiwinanda *et al.*, 2019), além de análises histopatológicas (Hlavcheva *et al.*, 2019). No desenvolvimento de fármacos, as RNCs são utilizadas em diversas aplicações, como para prever bioatividade (Wallach *et al.*, 2015), em estudos de interação alvo-ligante (Torng; Altman, 2019) e de interações proteína-proteína (Mallet *et al.*, 2022).

Capítulo 4

Metodologia

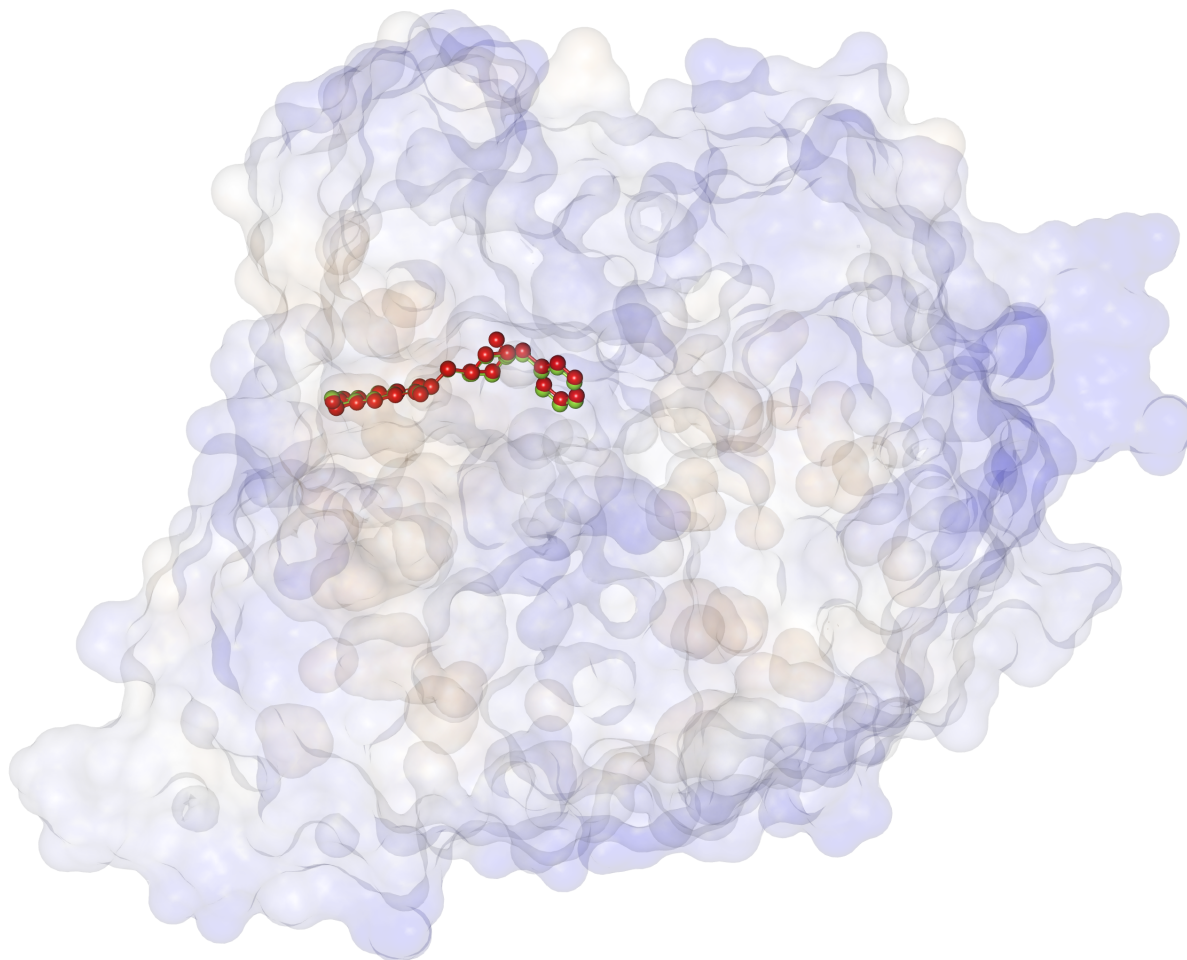
4.1 Protocolo de docking

Os cálculos de docking molecular foram realizados com a enzima hAChE, utilizando sua estrutura cristalográfica disponível no **Protein Data Bank** (PDB) sob o código 6O4W. Foram analisadas 32 moléculas, incluindo canabinóides, canabinomiméticos de origem natural, além do donepezil (DNP) como molécula de referência e a levodopa (LVD) como controle negativo. As simulações foram conduzidas por meio do Algoritmo Genético Lamarckiano (Morris *et al.*, 1998), com uma população de 300 soluções para a busca conformacional e um número máximo de avaliações de $2,5 \times 10^7$. Os valores de afinidade de ligação (binding scores) foram reportados em kcal/mol. Todos os cálculos foram performados com o complexo proteína-ligante rígido. Os estados de protonação dos resíduos de histidina na enzima foram manualmente configurados na suíte do AutoDock Tools, baseado nas predições feitas pelo software PROPKA (Søndergaard *et al.*, 2011; Olsson *et al.*, 2011).

O *grid box* foi definido com centro nas coordenadas: $x = 87,575$, $y = 84,670$ e $z = -4,677$, abrangendo um total de 307257 pontos de grade, com dimensões de 68 pontos em x , 68 em y e 72 em z . As simulações consideraram como sítio ativo os resíduos localizados na cavidade GORGE (do inglês para *Gorge of the Active Site*), uma fenda catalítica profunda e estreita situada na região central da enzima, conhecida por concentrar resíduos aromáticos, básicos e hidrofóbicos altamente conservados, fundamentais para o reconhecimento molecular, orientação do substrato até o sítio catalítico e estabilização de inibidores (De Paula *et al.*, 2009). Os cálculos de docking molecular foram conduzidos empregando ligantes parametrizados com cargas quânticas derivadas do método *Potential Grid-based - CHelpG* (*Charges from Electrostatic Potentials using a Grid based method*) (Breneman; Wiberg, 1990).

A Figura 4.1 apresenta o resultado do protocolo de redocking do ligante donepezil na estrutura 6O4W. A pose experimental (em verde-limão) foi comparada com a melhor pose prevista (em vermelho), correspondente à conformação protonada do DNP, que apresentou energia de afinidade de $-11,60$ kcal/mol. O redocking resultou em um RMSD de $0,26$ Å em relação à estrutura cristalográfica, validando a parametrização empregada no protocolo. A delimitação da cavidade GORGE como sítio ativo foi adotada como referência nas simulações, por representar a região de maior relevância para a alta afinidade entre os ligantes e a enzima hAChE.

Figura 4.1: Superposição da melhor pose do estudo de redocking.



4.2 Cálculos de estrutura eletrônica

Todas as moléculas foram otimizadas a fim de se obter a conformação correspondente ao mínimo global de energia, utilizando o software computacional Gaussian 16 (Frisch *et al.*, 2016). Foi empregado o funcional híbrido B3LYP, com a função de base 6-311+G(d, p) (Becke, 1992; Lee *et al.*, 1988; Vosko *et al.*, 1980; Stephens *et al.*, 1994). As otimizações foram realizadas tanto no vácuo quanto em meio com solvente implícito, sendo utilizada a água como solvente. O modelo implícito de solvatação adotado foi o SMD (do inglês *Solvation Model Based on Electron Density*) (Marenich *et al.*, 2009).

4.3 Coleta de dados ADMET e de estrutura eletrônica

Os descritores das propriedades de absorção, distribuição, metabolismo, excreção e toxicidade (ADMET) foram obtidas do software SwissADME (Daina *et al.*, 2017). Os descritores de estrutura eletrônica foram obtidos com o software Gaussian 16 (Frisch *et al.*, 2016) e com o software Multiwfn (Lu; Multiwfn, 2012). Os dados completos dos descritores de estrutura eletrônica e das propriedades ADMET obtidos neste trabalho podem ser consultados, respectivamente, nas Tabelas 7.1 e 7.2, localizadas no Apêndice.

4.4 Algoritmos desenvolvidos e implementados

4.4.1 Ferramentas usadas

Todos os algoritmos foram desenvolvidos usando linguagem de programação Python, na versão 3.11. Diversas bibliotecas de machine learning e ciência de dados foram usadas, como NumPy (Harris *et al.*, 2020), Pandas (team, 2020; McKinney, 2010), Matplotlib (Hunter, 2007), scikit-learn (Pedregosa *et al.*,

2011), TensorFlow (Martín Abadi *et al.*, 2015) e PyTorch (Paszke *et al.*, 2019) para processamento, análise, modelagem dos dados e construção dos modelos de redes neurais artificiais.

4.4.2 Software HeatLQC

O algoritmo de similaridade desenvolvido para o software HeatLQC tem por objetivo avaliar a similaridade entre diferentes fármacos a partir de um conjunto de dados, indicando classes de fármacos, bem como servindo de ferramenta para estudos de reposicionamento de fármacos. Inicialmente o algoritmo reconhece cada fármaco como um vetor compostos por seus descritores (propriedades ADMET e de estrutura eletrônica).

$$\begin{aligned} S &= \{A, B, C, \dots, N\}, \quad A = \text{Mol0}, B = \text{Mol1}, \dots, N - \text{ésima molécula}, \\ A &= \{x_1, x_2, \dots, x_M\}, \quad x_1 = \text{Log P}, x_2 = \text{HOMO}, \dots, M - \text{ésima propriedade} \end{aligned} \quad (4.1)$$

em seguida, esse conjunto de dados de entrada S é autoescalonado, gerando um novo conjunto S' normalizado.

$$S' = \{A', B', C', \dots, N'\}, \quad A' = z_1, z_2, \dots, z_M, \quad z_i = \frac{x_i - \mu}{\sigma} \quad (4.2)$$

onde μ é a média e σ o desvio padrão. O próximo passo do algoritmo consiste em calcular os coeficientes de correlação produto-momento, ρ , também denominados de coeficientes de Pearson, que variam entre +1 e -1.

$$\rho_{A', B'} = \frac{\text{cov}(A', B')}{\sigma_{A'} \sigma_{B'}} \approx \frac{\sum_{i=1}^M (z_i^{A'} - \bar{z}_{A'}) (z_i^{B'} - \bar{z}_{B'})}{\sqrt{\sum_{i=1}^M (z_i^{A'} - \bar{z}_{A'})^2} \sqrt{\sum_{i=1}^M (z_i^{B'} - \bar{z}_{B'})^2}} \quad (4.3)$$

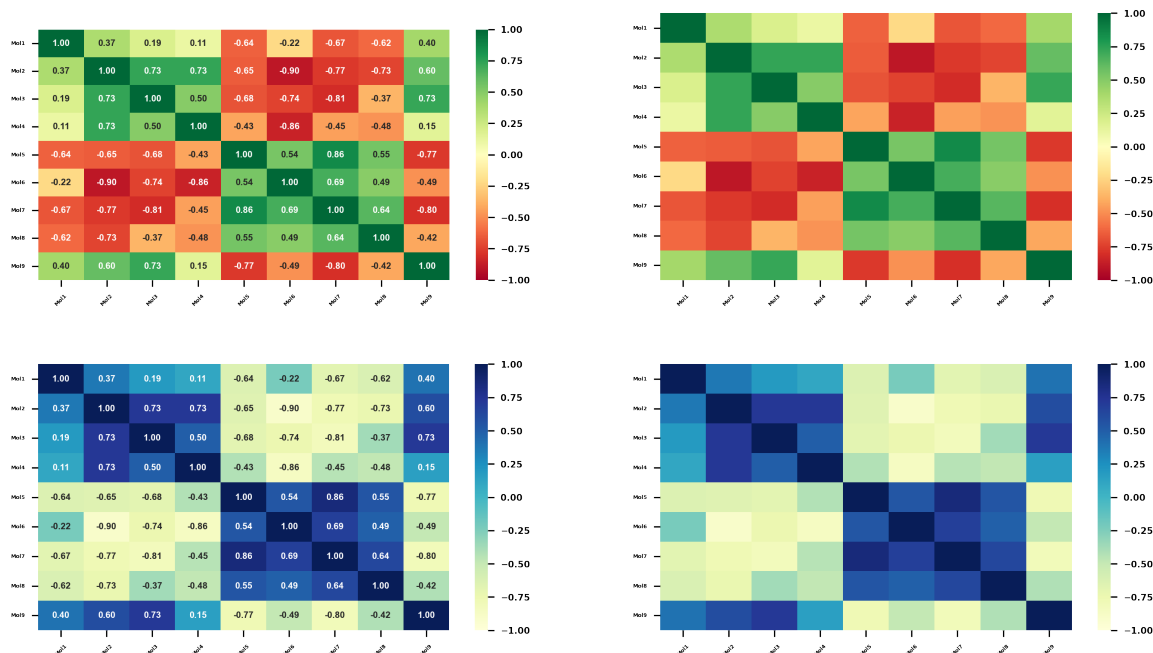
por fim, os coeficientes ρ são dispostos em uma matriz de correlação, M_{corr} , representada visualmente como um mapa de calor (heatmap).

$$M_{\text{corr}} = \begin{bmatrix} \rho_{A'A'} & \rho_{A'B'} & \rho_{A'C'} & \cdots & \rho_{A'N'} \\ \rho_{B'A'} & \rho_{B'B'} & \rho_{B'C'} & \cdots & \rho_{B'N'} \\ \rho_{C'A'} & \rho_{C'B'} & \rho_{C'C'} & \cdots & \rho_{C'N'} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{N'A'} & \rho_{N'B'} & \rho_{N'C'} & \cdots & \rho_{N'N'} \end{bmatrix} \quad (4.4)$$

Algoritmo 1 Algoritmo de matriz de similaridade

$D \leftarrow$ dados de entrada, D = objeto de dados estruturado com linhas e colunas
 $SD \leftarrow S(D)$, S = operador de normalização (Scaler)
 $SDT \leftarrow T(SD)$, T = operador de transposição
 $C \leftarrow D(\text{colunas})$
 $ND \leftarrow \{SDT, C\}$, ND = novo objeto de dados estruturado
 $M_{\text{corr}} \leftarrow MC(ND)$, MC = operador de matriz de correlação

Figura 4.2: Exemplos de heatmaps gerados pelo programa desenvolvido.



4.4.3 PCA Analytics

O software PCA Analytics tem como objetivo oferecer uma ferramenta de análise molecular voltada para a investigação de descritores ADMET e de estrutura eletrônica e do perfil de similaridade entre moléculas. Seu uso é especialmente relevante em aplicações como triagem virtual e reposicionamento de fármacos, permitindo a identificação de compostos estruturalmente ou funcionalmente semelhantes a moléculas bioativas conhecidas.

O programa utiliza o formalismo dos algoritmos de machine learning PCA e k-means discutidos na seção de métodos de data science. Os dados são automaticamente centrados pelo algoritmo PCA da biblioteca scikit-learn, subtraindo-se a média de cada variável. Todavia, utilizamos a saída do algoritmo desenvolvido no programa HeatLQC, que consiste na matriz de correlação M_{corr} , para gerar o espaço de característica onde o algoritmo k-means atua. Para tanto, utilizamos o algoritmo clássico de Lloyd, mas colocando como entrada a matriz M_{corr} gerada pelo algoritmo de similaridade desenvolvido para o software HeatLQC.

Algoritmo 2 Algoritmo de Lloyd com matriz de similaridade como entrada

$M_{corr} \leftarrow$ dados de entrada

Inicialização: centróides aleatórios $C = \{c^1, \dots, c^k\}$

Repetir para $i = 1, 2, \dots, k$

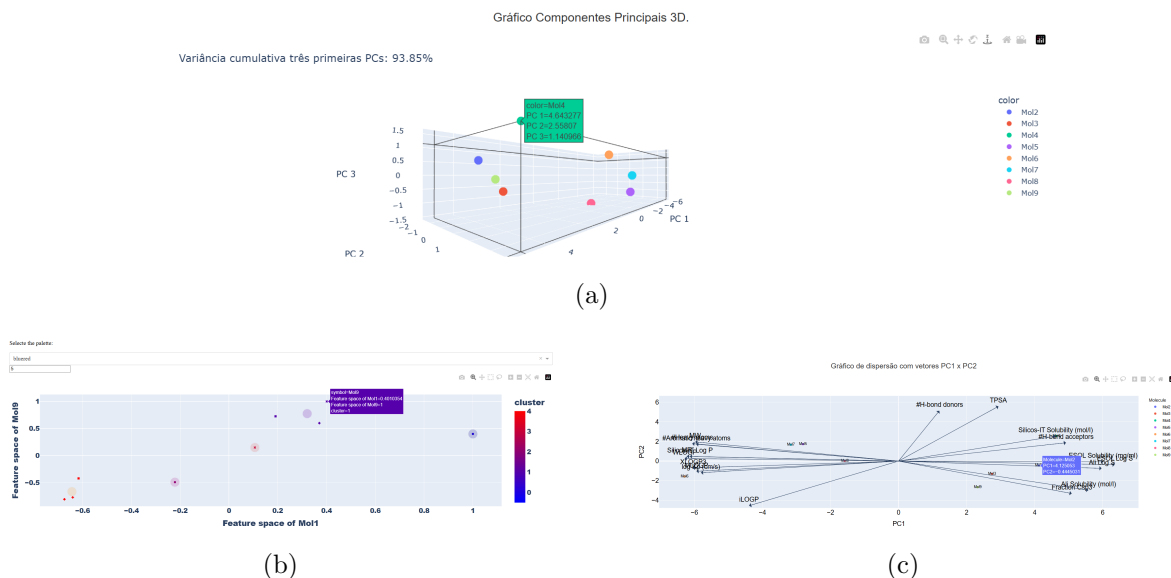
$b(x) = \arg \min_i \{\Delta(x, c_i)\}, \quad \forall x \in M_{corr}$

Atualizar todos os centróides dos clusters em C :

$$c^i = \frac{1}{|A^i|} \sum_{x \in A^i} x, \quad \forall i \in \{1, 2, \dots, k\}$$

Até que a média dos centróides dos clusters pare de mudar

Figura 4.3: Exemplo de funcionalidade do software PCA Analytics. (a) Visualização tridimensional para as três primeiras componentes da PCA; (b) Resultado do algoritmo k-means no espaço de características gerado a partir da matriz M_{corr} ; (c) Gráfico de dispersão bidimensional das amostras no espaço definido por duas componentes principais selecionadas pelo usuário, com vetores de *loadings* representando a contribuição de cada propriedade molecular para essas componentes. A direção e o comprimento dos vetores indicam a influência relativa de cada variável original (como solubilidade, logP, TPSA, entre outras) na formação dos eixos principais. As moléculas estão posicionadas conforme suas projeções nas componentes selecionadas, permitindo visualizar padrões de agrupamento e correlação com as propriedades descritoras.

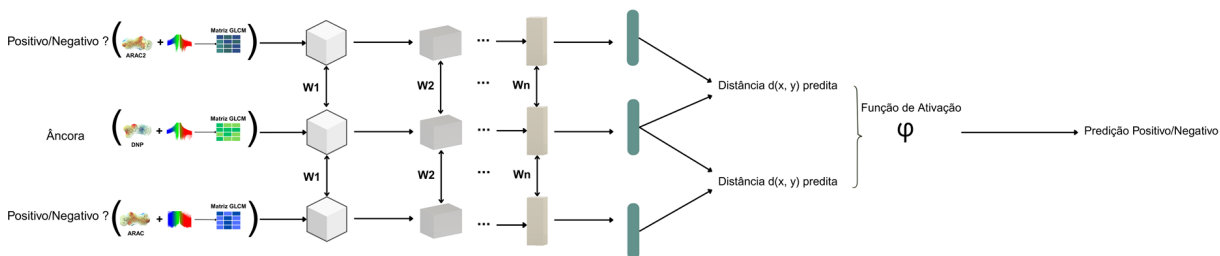


4.4.4 Rede neural siamesa para similaridade via Imagens MEP

Arquitetura do Modelo

O modelo proposto tem por objetivo prever a similaridade molecular a partir de representações visuais e textuais extraídas de imagens MEP dos compostos estudados e das matrizes GLCM extraídas a partir dos diagramas RDG. O modelo possui uma arquitetura do tipo siamesa, na qual duas entradas paralelas — a âncora (a) e o par (positivo (p) ou negativo (n)) — são processadas por redes neurais artificiais idênticas que compartilham pesos (Figura 4.4).

Figura 4.4: Esquema ilustrativo da rede siamesa para predição de similaridade molecular desenvolvida. Apesar da pré-definição inicial de a, p e n, o modelo aprende as distâncias entre as imagens MEP e diagramas RDG durante o treinamento, retornando as distâncias mais próximas entre as imagens



O processamento das imagens MEP é feito por uma rede do tipo RNC, composta por três blocos sequenciais do tipo Conv2D → ReLU → Pooling → AdaptiveAvgPool, seguidos por uma camada linear de projeção que gera os vetores *embedding*, contendo cada 128 dimensões. Paralelamente, as informações textuais das imagens RDG, são processadas por um perceptron multicamada (do inglês *multilayer perceptron*, MLP), com duas camadas lineares, responsável por transformar as matrizes GLCM (vetores de dimensão 64) em *embeddings* de dimensão 64. As saídas da RNC e do MLP são então concatenadas, formando um vetor de 192 dimensões que é então porjetado em um espaço latente comum de 128 dimensões por meio de um projetor multimodal composto por duas camadas lineares intercaladas com funções de

ativação ReLU e *Dropout*, sendo o hiperpâmetro probabilidade da função *Dropout* definido como igual a 0.3.

O treinamento é conduzido utilizando a função de perda tripla, que promove a aproximação dos embeddings de moléculas semelhantes e o distanciamento das moléculas dissimilares. Para definição inicial dos tripletos utilizou-se a similaridade de Tanimoto calculada a partir dos fingerprints SMILES. Para avaliar o modelo utilizou-se a métrica acurácia, que mede a previsão de previsões corretas em relação ao total de previsões

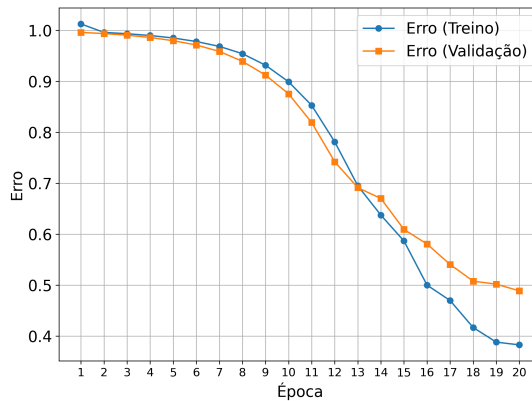
$$\text{Acurácia} = \frac{\text{Soma das previsões corretas}}{\text{Soma total de previsões}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (4.5)$$

onde TP, TN, FP, FN são as siglas em inglês: *true positive*, *true negative*, *false positive* e *false negative* respectivamente. Outras formas pelas quais o modelo foi avaliado foram por meio da ferramenta Grad-CAM, que demonstra as regiões da imagem que o modelo deteve maior atenção para predição da similaridade, bem como da comparação entre as matrizes de correlação geradas via coeficiente de tanimoto S_T e distância invertida, d^{-1} predita pelo modelo. A distância invertida é definida como

$$d^{-1} = \frac{1}{1 + d(x, y)} \quad (4.6)$$

onde x e y são as duas imagens que se quer comparar e $d(x, y)$ é a distância predita pelo modelo. O algoritmo 3 descreve o modelo de forma concisa

Figura 4.5: Erro medido com métrica Acurácia durante treinamento para os dados de treino e validação



Algoritmo 3 Treinamento da rede siamesa para similaridade molecular via MEP e GLCM

Require: Conjunto de imagens MEP e vetores GLCM, matriz de similaridade S_T , margem α , número de épocas E e modelo Φ

Inicializar CNN, MLP e projetor multimodal de Φ

for e = 1 **to** E **do**

▷ Para cada época

Construir tripletos (a, p, n) com:

$S_T(a, p) \geq 0.70$ (positivo) e $S_T(a, n) \leq 0.40$ (negativo)

▷ Ou outro critério específico

for cada tripleto (a, p, n) **do**

$z_a \leftarrow \Phi(a_{MEP}, a_{glcm})$

$z_p \leftarrow \Phi(p_{MEP}, p_{glcm})$

$z_n \leftarrow \Phi(n_{MEP}, n_{glcm})$

Calcular perda:

$\mathcal{L} \leftarrow \max(0, \|z_a - z_p\|^2 - \|z_a - z_n\|^2 + \alpha)$

Atualizar os pesos via backpropagation ($\mathcal{L}$)

end for

end for

return Φ

4.4.5 Rede neural generativa para geração de moléculas-VAELSTM

Dataset

Foi utilizado um dataset com um total de 310253 SMILES de moléculas, obtidas a partir dos bancos de dados públicos ZINC (Irwin; Shoichet, 2005), GBD (Fink *et al.*, 2005; Fink; Reymond, 2007), e de um terceiro banco de dados privado, compilado por nosso grupo, contendo SMILES de canabinóides e canabinomiméticos de origem natural. As quantidades de SMILES para cada conjunto foram: 250000 do ZINC, 60000 do GBD e 253 do dataset privado desenvolvido por nossa equipe. Não houve um critério específico de escolha para esses conjuntos públicos além da facilidade de acesso e da boa reprodutibilidade.

Posteriormente, aprimoramos a coleta de SMILES por meio de requisições HTTP em tempo real à API do banco de dados ZINC, visando obter moléculas de classes específicas de interesse. Esse segundo processo resultou em um conjunto consideravelmente maior, na ordem de milhões de SMILES, e está sendo aprofundado em um trabalho em desenvolvimento que será submetido como artigo científico.

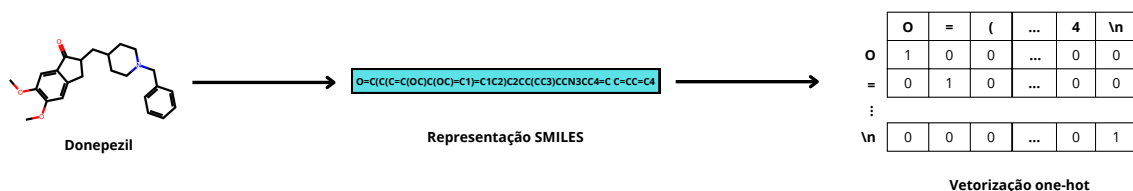
Arquitetura do Modelo desenvolvido

O algoritmo proposto tem por objetivo gerar sequências de SMILES válidas a partir de amostras no espaço latente gerado por um VAE. O modelo consiste em um VAE que utiliza células LSTM como decodificador. Inicialmente, o codificador VAE e o decodificador LSTM são carregados, juntamente com um módulo de mapeamento, f_{ϕ}^{map} . Em seguida, define-se um vocabulário Σ , contendo os tokens utilizados na geração, incluindo os marcadores especiais de início (“!”) e de término (“E”) da sequência.

Para cada uma das n tentativas de geração, amostra-se um vetor latente $z \in \mathbb{R}^{d_z}$ — onde d_z é a dimensão do espaço latente — a partir de uma distribuição normal $\mathcal{N}(0, I_{d_z})$, com ruído uniforme adicional $\mathcal{U}(-0.5, 0.5)$ incorporado em versões posteriores do modelo, com o intuito de promover maior diversidade no processo generativo. Esse vetor z é transformado nos estados ocultos iniciais h_0 e c_0 das células LSTM por meio do mapeamento do módulo f_{ϕ}^{map} , que condiciona a geração à amostra latente.

A sequência é então iniciada com o token de partida “!”, que é convertido para codificação do tipo *one-hot* (Figura 4.6) e utilizado como primeira entrada para o decodificador LSTM. A cada passo de tempo t , a célula LSTM gera uma saída o_t , da qual se obtém uma distribuição de probabilidade p_t sobre os próximos tokens, por meio de uma função de ativação softmax com temperatura τ , permitindo controle sobre a aleatoriedade — ou diversidade — das saídas no processo generativo.

Figura 4.6: Representação esquemática da codificação do tipo *one-hot*



O próximo token é escolhido como aquele que possui a maior probabilidade dentro da distribuição p_t , e o processo continua até que o token de parada “E” seja emitido ou o número máximo de passos T_{max}

seja atingido. Ao final de cada tentativa, a sequência s gerada é verificada quanto à validade sintática como uma string SMILES. Apenas as sequências válidas são adicionadas ao conjunto $\mathcal{S}$ de resultados, o qual é retornado ao término do processo.

Algoritmo 4 Versão concisa do algoritmo desenvolvido

Require: Número de tentativas $n > 0$, Dimensão latente $d_z > 0$

Ensure: Conjunto $\mathcal{S}$ de strings SMILES válidas

```

1: Inicializar modelos VAE, LSTM e mapeamento ( $f_\phi^{map}$ )
2: Definir vocabulário  $\Sigma$  com tokens de início ("!") e fim ("E")
3:  $\mathcal{S} \leftarrow \emptyset$  ▷ Conjunto de resultados
4: for  $i \leftarrow 1$  to  $n$  do
5:   Amostrar  $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_{d_z}) + \mathcal{U}(-0.5, 0.5)$ 
6:   Inicializar estados ocultos:  $(\mathbf{h}_0, \mathbf{c}_0) \leftarrow f_\phi^{map}(\mathbf{z})$ 
7:   Iniciar sequência:  $s \leftarrow "!", \mathbf{x}_0 \leftarrow \text{one-hot}("!")$ 
8:   for  $t \leftarrow 1$  to  $T_{\max}$  do
9:     Calcular saída:  $\mathbf{o}_t \leftarrow f_\phi^{\text{LSTM}}(\mathbf{x}_{t-1}, \mathbf{h}_{t-1}, \mathbf{c}_{t-1})$ 
10:    Obter distribuição:  $\mathbf{p}_t \leftarrow \text{softmax}_\tau(\mathbf{o}_t)$ 
11:    Selecionar token:  $k \leftarrow \arg \max(\mathbf{p}_t)$ 
12:    if  $\Sigma[k] = "E"$  then break
13:    end if
14:    Atualizar sequência:  $s \leftarrow s \parallel \Sigma[k]$ 
15:    Preparar próxima entrada:  $\mathbf{x}_t \leftarrow \text{one-hot}(\Sigma[k])$ 
16:  end for
17:  if  $s$  is SMILES válido then  $\mathcal{S} \leftarrow \mathcal{S} \cup \{s\}$ 
18:  end if
19: end for
20: return  $\mathcal{S}$ 

```

Para avaliar o desempenho do modelo, foi calculada a variância explicada pelas duas primeiras componentes principais do espaço latente gerado em cada valor de temperatura. Adicionalmente, foram computadas as métricas de validade (Equação 4.7) e unicidade (Equação 4.8), que fornecem medidas quantitativas da correção sintática e da diversidade das moléculas geradas, respectivamente. A validade é definida como a razão entre o número de moléculas válidas geradas ($N_{\text{válidas}}$) e o número total de tentativas de geração (N_{total}). Já a unicidade corresponde à proporção de moléculas únicas ($N_{\text{únicas}}$) em relação ao total de moléculas válidas geradas ($N_{\text{válidas}}$).

$$\text{validade} = \frac{N_{\text{válidas}}}{N_{\text{total}}} \quad (4.7)$$

$$\text{unicidade} = \frac{N_{\text{únicas}}}{N_{\text{válidas}}} \quad (4.8)$$

Capítulo 5

Resultados e Discussão

5.1 Docking e estrutura eletrônica

A análise das interações moleculares no sítio catalítico da hAChE revelou, inicialmente, o perfil do complexo hAChE-DNP, molécula de referência neste estudo. O DNP apresentou uma energia de interação de -11,60 kcal/mol, refletindo sua capacidade de ancoragem no sítio ativo GORGE da enzima. Esse desempenho decorreu, principalmente, da presença de um grupamento amina protonado, que permitiu a formação de interações cátion-pi com os resíduos aromáticos W86 (4,26 Å) e Y337 (3,75 Å). Além dessas, o DNP estabeleceu uma ligação de hidrogênio não convencional com o resíduo H447 (3,60 Å) e uma interação pi-sigma com P338 (3,88 Å), compondo um perfil de reconhecimento molecular característico de inibidores da hAChE (Figura 5.1). Do ponto de vista eletrônico, o DNP apresentou orbitais de fronteira com valores de LUMO+1 em -0,89 eV, LUMO em -1,74 eV e HOMO em -6,15 eV, indicando equilíbrio entre estabilidade e reatividade. Esse perfil foi complementado pelos índices condensados de densidade dual, CDD^+ de 0,146 e CDD^- de -0,084, os quais refletiram uma distribuição eletrônica favorável para interações nucleofílicas (Figura 5.2).

Figura 5.1: Interações do DNP com alguns dos principais resíduos da hAChE

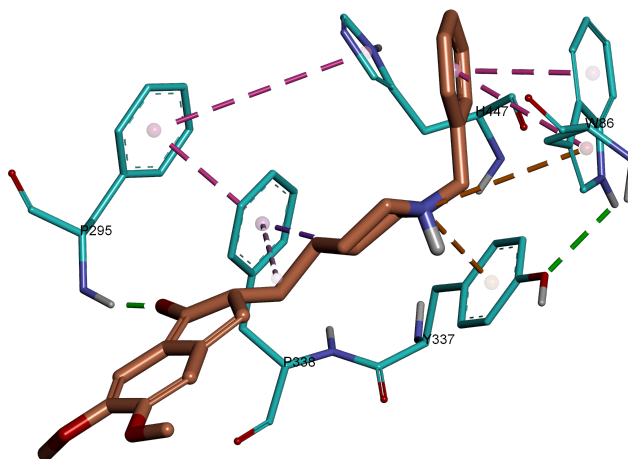
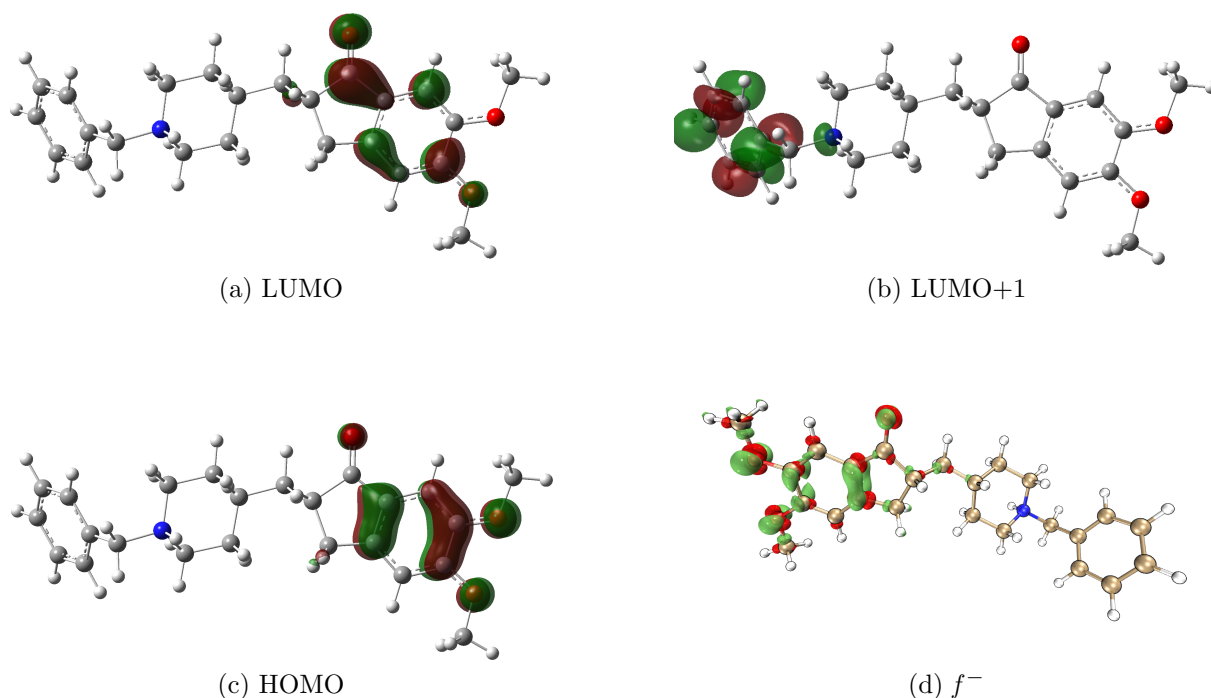


Figura 5.2: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o DNP.



De forma geral, apesar da ausência de átomos com baixa densidade eletrônica nas moléculas estudadas (canabinóides, canabinóides raros e canabinomiméticos), sete moléculas do conjunto testado foram capazes de mimetizar o padrão de interações observado para o DNP. Esses compostos apresentaram valores de afinidade compatíveis e interações relevantes com os resíduos W86, Y337, H447 e P338. Nesse contexto, destacou-se a molécula PRT, que obteve uma energia de interação de -10,32 kcal/mol, o mais próximo do DNP entre todas as moléculas avaliadas.

O desempenho de PRT não se limitou ao valor da energia de interação, tendo apresentado perfil de interação semelhante ao do DNP (Figura 5.3). As interações de destaque do DNP no GORGE foram: interação pi-pi do tipo stacked com W86 (3,86 Å) e duas interações pi-alquil, uma com Y337 (4,40 Å) e outra com P338 (5,08 Å). Também apresentou uma interação de van der Waals com H447 (6,00 Å). As distâncias observadas permaneceram dentro de uma faixa que favoreceu a estabilização do complexo. A comparação dos parâmetros eletrônicos reforçou essa proximidade, com valores de LUMO+1 em -0,42 eV, LUMO em -0,50 eV e HOMO em -6,00 eV. Da mesma forma, os índices CDD^+ (0,098) e CDD^- (-0,087) corroboraram a capacidade da molécula de ocupar, tanto espacial quanto eletronicamente, a cavidade da enzima de maneira semelhante ao DNP (Figura 5.4).

Figura 5.3: Interações do PRT com os resíduos da GORGE, mimetizando a interação cátion-pi característica do DNP.

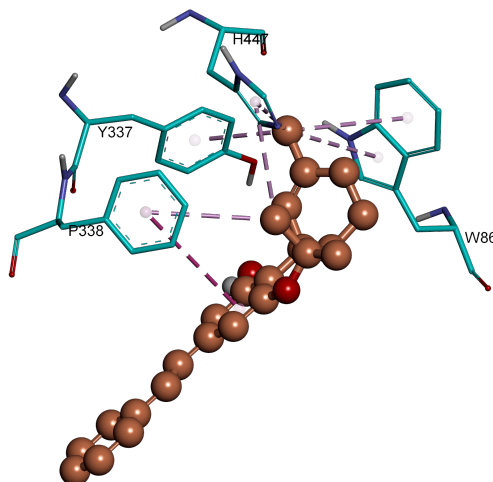
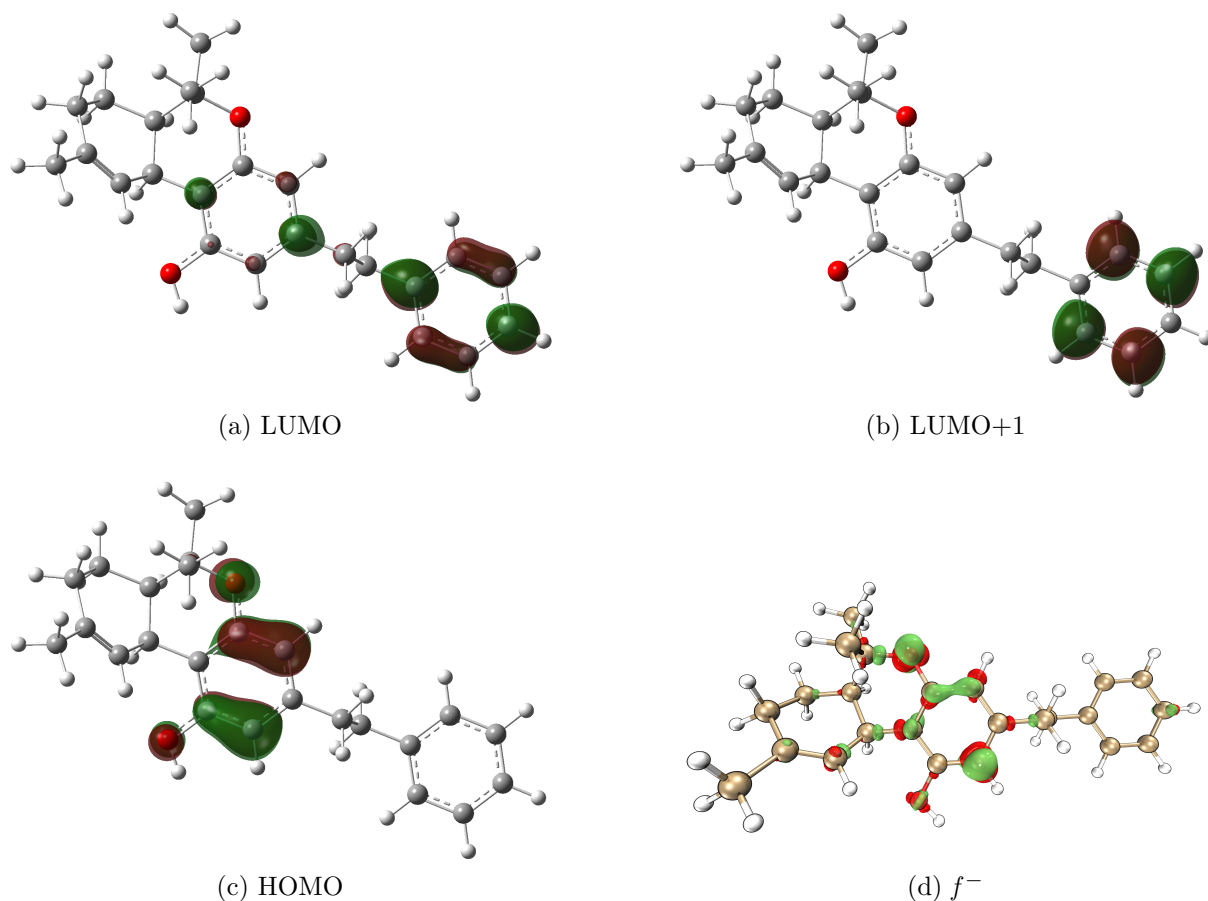


Figura 5.4: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o PRT.



A molécula DMDD, apresentou energia de interação de -9,70 kcal/mol na formação do complexo com a enzima, tendo estabelecido interações pi-sigma com W86 (3,71 Å) e Y337 (3,44 Å), além de uma interação pi-alkil com P338 (4,44 Å) e uma interação pi-pi T-shaped com H447 (4,78 Å). Suas principais interações podem ser visualizadas na Figura 5.5. Seus orbitais apresentaram valores de LUMO+1 em -0,48 eV, LUMO em -1,87 eV e HOMO em -6,09 eV, enquanto os índices CDD^+ (0,098) e CDD^- (-0,068) indicaram um perfil eletrônico comparável, embora mais deslocado em direção a estados eletrônicos mais estabilizados, quando comparado ao DNP (Figura 5.6).

Figura 5.5: Interações do DMDD com resíduos da GORGE

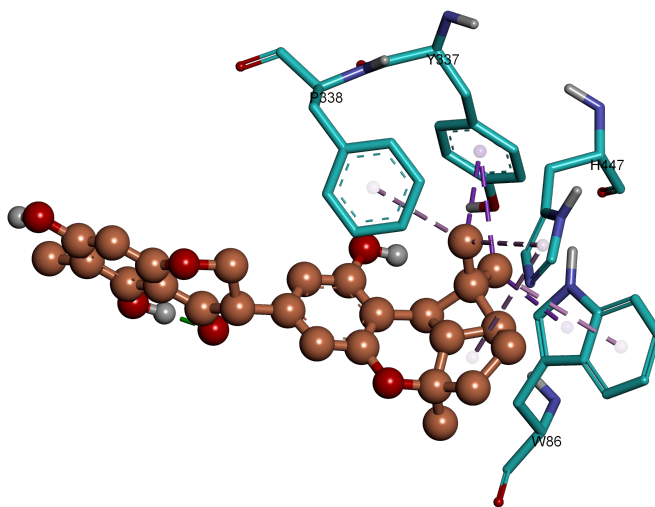
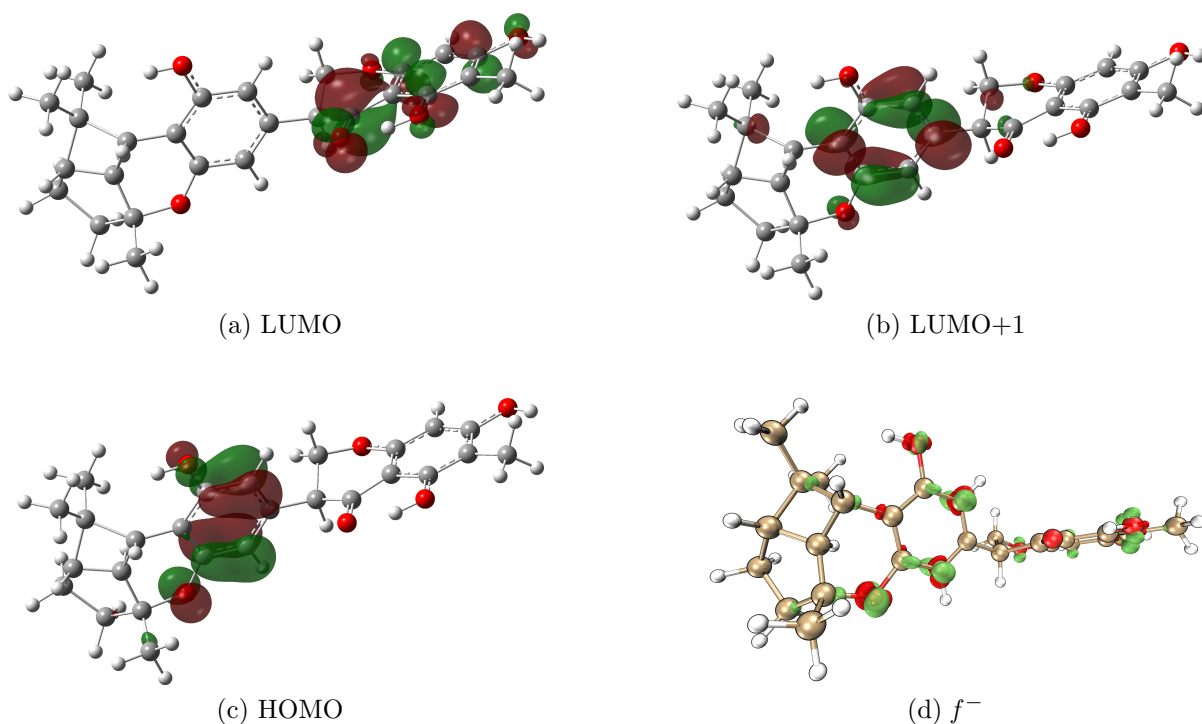


Figura 5.6: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o DMDD.



A molécula MCRA (ver Figura 5.7) apresentou energia de interação de -9,48 kcal/mol. Suas interações incluíram pi-pi T-shaped com W86 (5,30 Å), pi-sigma com Y337 (3,58 Å), pi-pi stacked com P338 (5,34 Å) e pi-pi T-shaped com H447 (4,30 Å). Os valores dos orbitais foram LUMO+1 em -0,48 eV, LUMO em -1,70 eV e HOMO em -5,69 eV, e os índices CDD^+ (0,024) e CDD^- (-0,030), que podem ser visualizados na Figura 5.8, foram consistentes com a presença de regiões capazes de participar de interações não covalentes.

Figura 5.7: Interações do MCRA com resíduos da GORGE

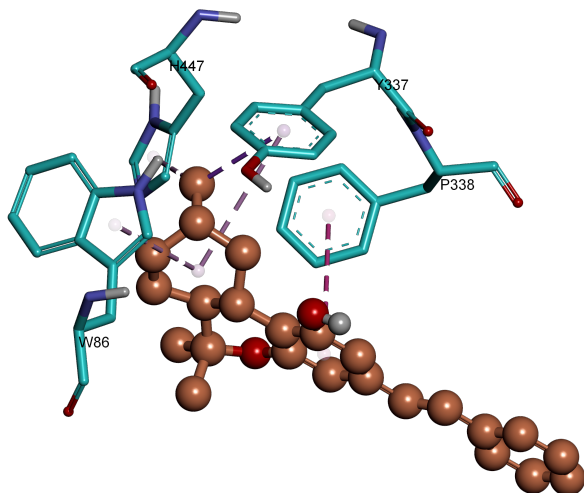
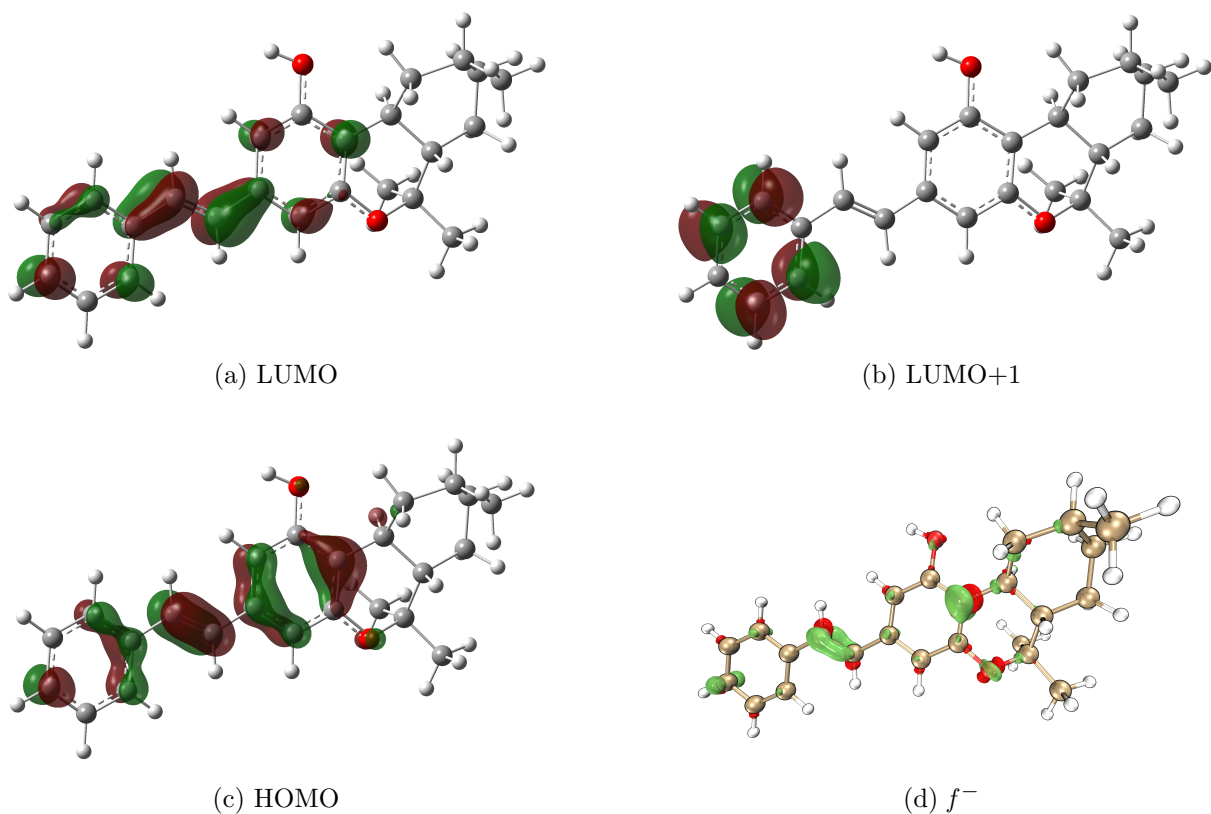


Figura 5.8: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o MCRA.



DMDE obteve energia de interação de -9,31 kcal/mol, com interações pi-pi T-shaped com W86 (4,84 Å) e Y337 (3,82 Å), pi-pi stacked com P338 (5,27 Å) e uma interação de van der Waals com H447 (6,00 Å). As principais interações do DMDE com os resíduos de aminoácido da GORGE podem ser vistas na Figura 5.9. Em relação aos resultados de estrutura eletrônica, que podem ser visualizados na Figura 5.10, seus orbitais apresentaram valores de LUMO+1 em -0,39 eV, LUMO em -1,86 eV e HOMO em -6,07 eV, com índices CDD^+ (0,097) e CDD^- (-0,070), dentro da faixa observada nas moléculas do conjunto.

Figura 5.9: Interações do DMDE com resíduos da GORGE

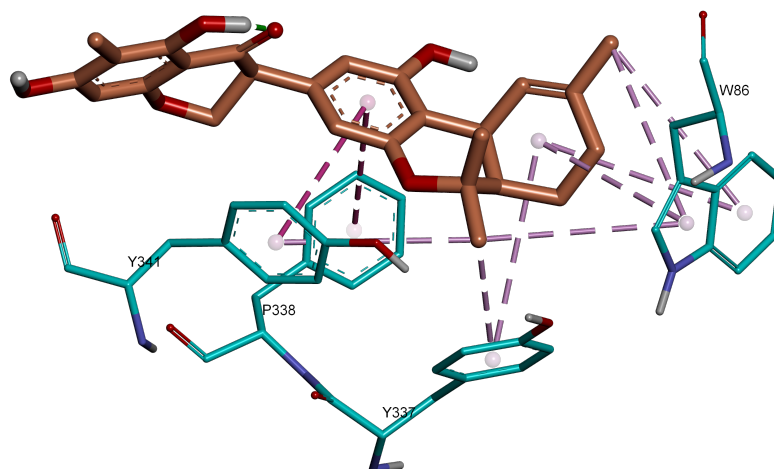
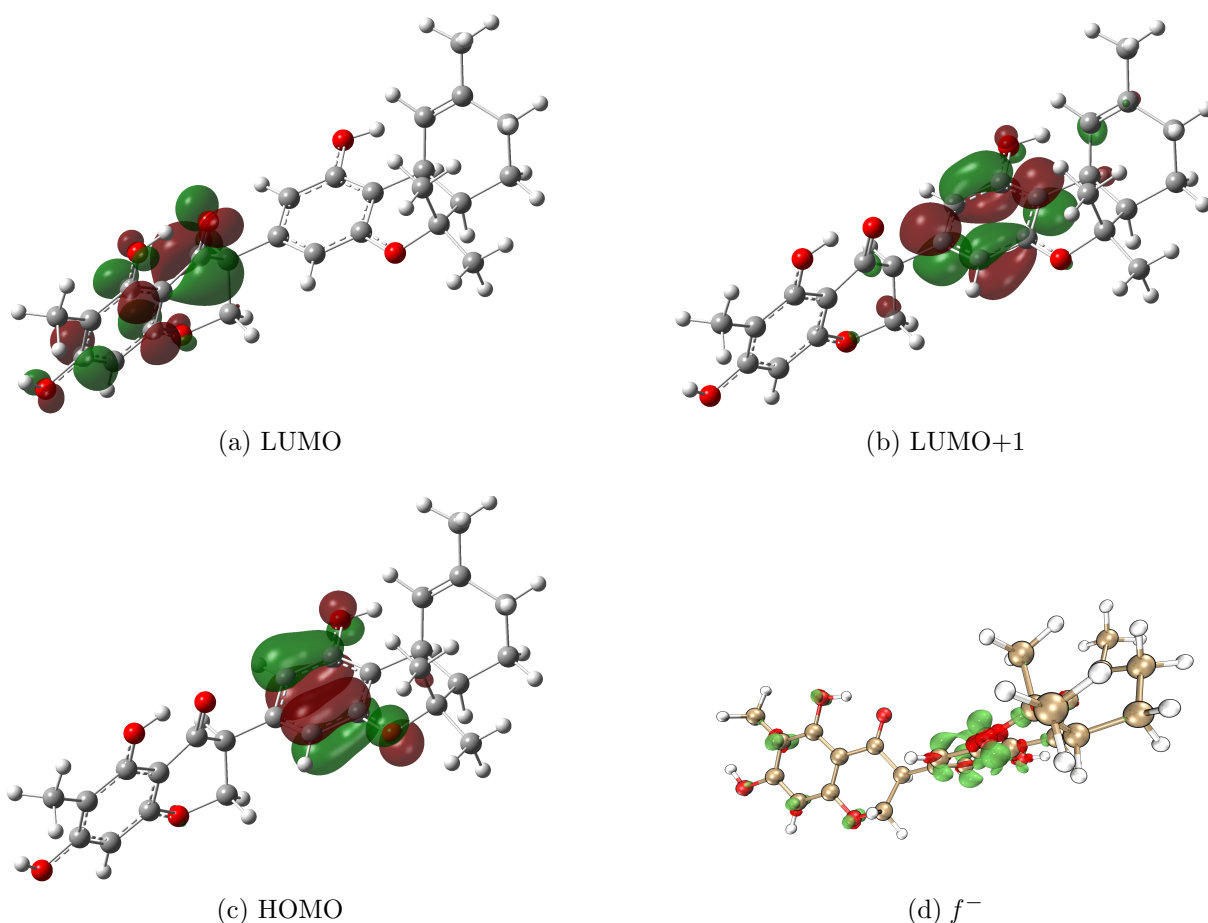


Figura 5.10: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o DMDE.



A Figura 5.11 mostra as principais interações entre o MCDC na GORGE da hAChE. A molécula MCDC teve energia de interação de -8,82 kcal/mol, tendo apresentado interações pi-sigma com W86 (3,50 Å) e Y337 (3,24 Å), pi-pi T-shaped com P338 (5,01 Å) e pi-alquil com H447 (4,96 Å). Seus orbitais apresentaram valores de LUMO+1 em -0,31 eV, LUMO em -1,55 eV e HOMO em -5,78 eV, enquanto os índices CDD^+ (0,014) e CDD^- (-0,015), embora com menor similaridade aos valores obtidos para o DNP, estiveram em concordância com uma distribuição eletrônica que favoreceu interações pi e hidrofóbicas (Figura 5.12).

Figura 5.11: Interações do MCDC com resíduos da GORGE

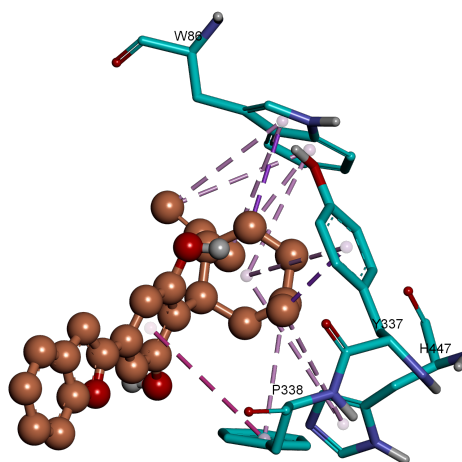
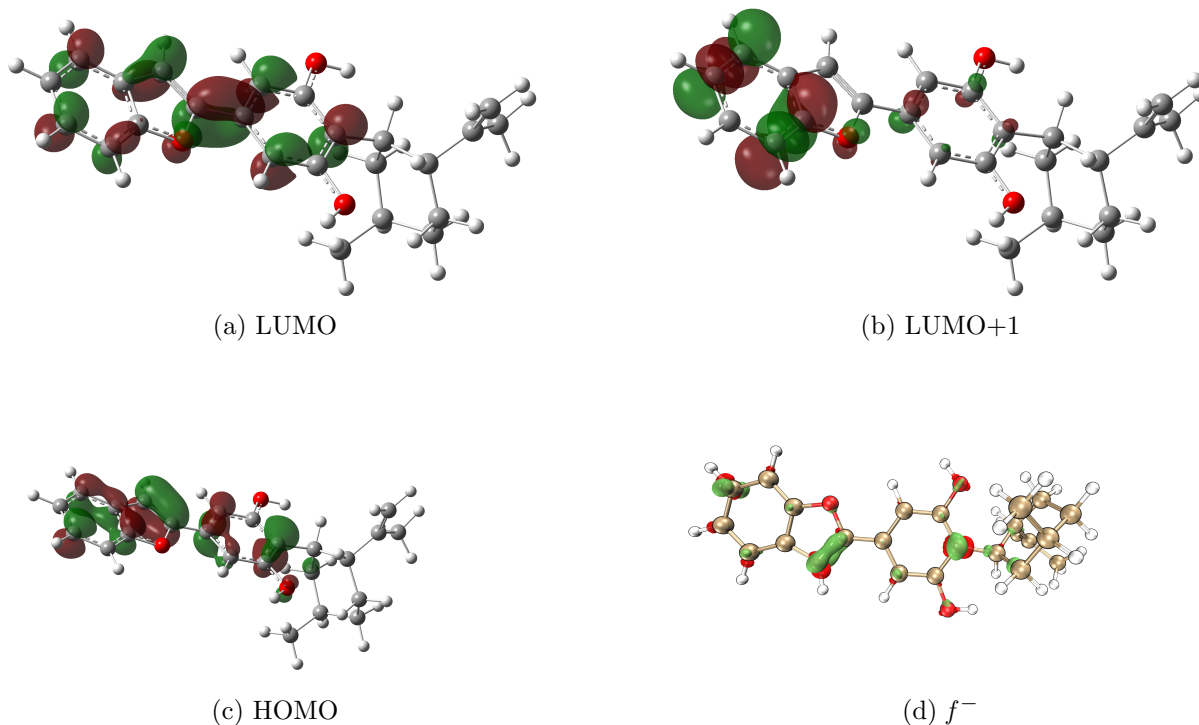


Figura 5.12: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o MCDC.



OMCNB apresentou energia de interação de -8,72 kcal/mol, estabelecendo interações (Figura 5.13) do tipo pi-pi T-shaped com W86 (4,62 Å) e interações pi-álquil com Y337 (5,02 Å), P338 (5,35 Å) e H447 (4,11 Å). Em relação aos resultados de estrutura eletrônica (Figura 5.14), os orbitais de fronteira analisados apresentaram valores de LUMO+1 em -0,40 eV, LUMO em -1,23 eV e HOMO em -5,91 eV, enquanto os índices CDD^+ (0,017) e CDD^- (-0,033) permaneceram dentro da faixa observada para o MCDC.

Figura 5.13: Interações do OMCNB com resíduos da GORGE

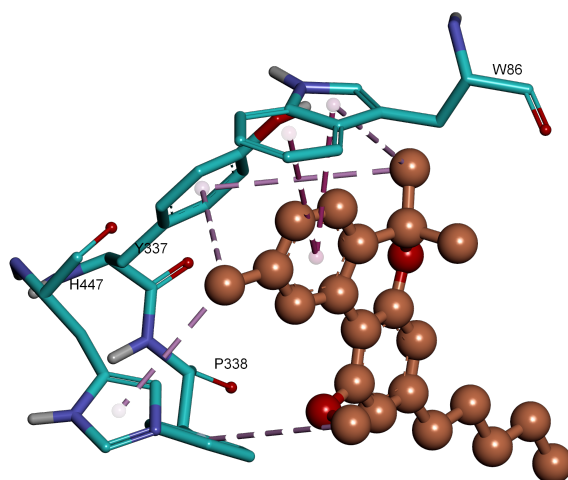
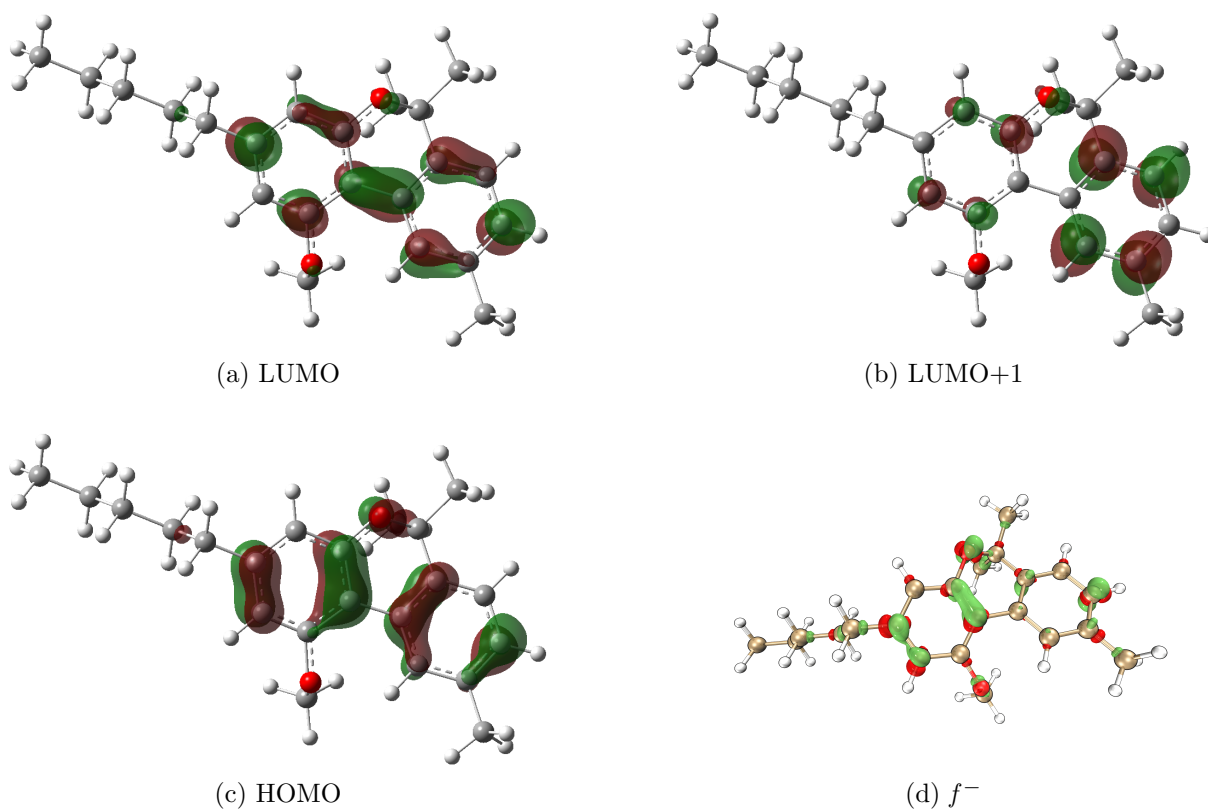


Figura 5.14: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o OMCNB.



Por fim, dentro os sete melhores resultados, o D9CT, com energia de interação de -8,61 kcal/mol, estabeleceu interações pi-sigma com W86 (4,13 Å), pi-alkil com Y337 (6,45 Å), pi-pi T-shaped com P338 (4,92 Å) e pi-alkil com H447 (4,41 Å) (Figura 5.15). Seu perfil eletrônico (Figura 5.16) apresentou valores de LUMO+1 em 0,03 eV, LUMO em -0,17 eV e HOMO em -5,97 eV, enquanto os índices CDD^+ (0,138) e CDD^- (-0,087) indicaram uma distribuição eletrônica com considerável similaridade em relação aos valores obtidos para o DNP.

Figura 5.15: Interações do D9CT com resíduos da GORGE

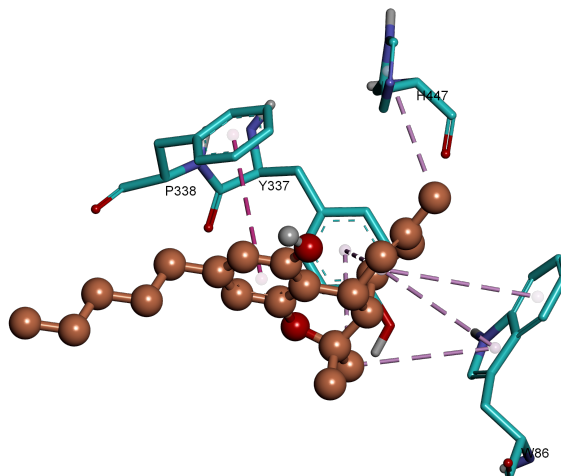
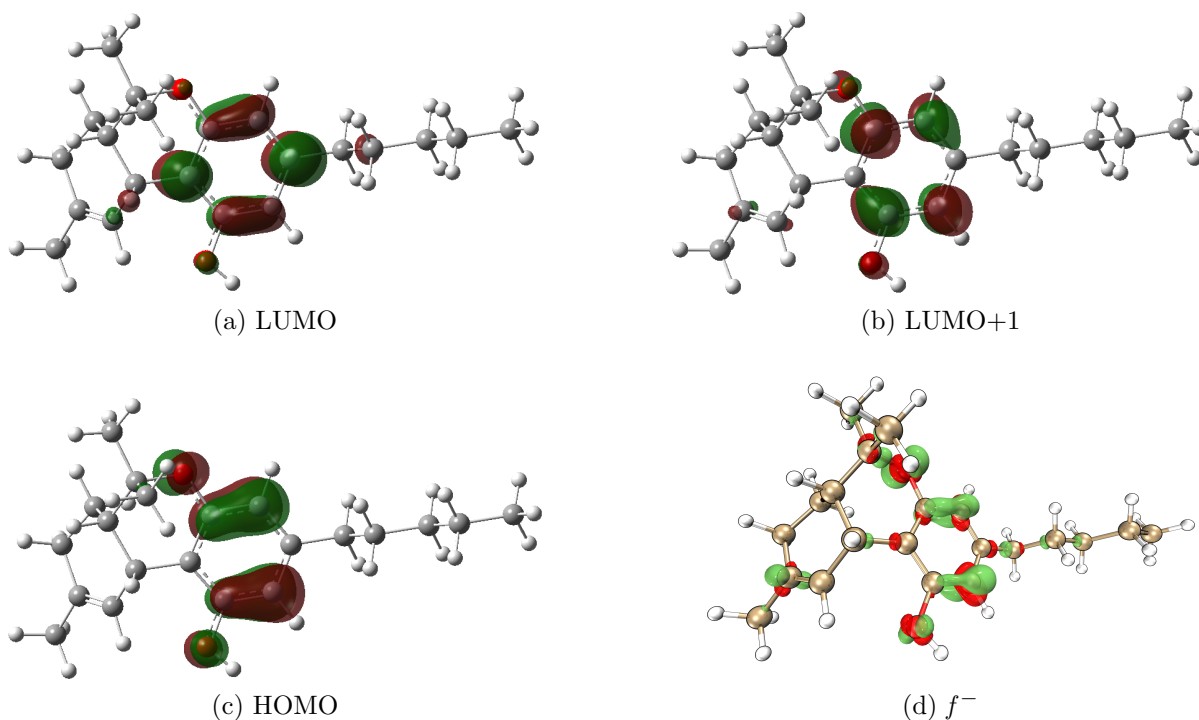


Figura 5.16: Representações dos orbitais LUMO, LUMO+1, HOMO e da função f^- para o D9CT.



A redução na densidade eletrônica, observada principalmente nos orbitais LUMO e nos índices CDD^+ e CDD^- , foi coerente com o aumento nas distâncias das interações que mimetizaram as interações cátion-pi do DNP com W86 e Y337. Ainda assim, PRT manteve distâncias para os resíduos W86, Y337 e P338 compatíveis com a faixa necessária para sustentar interações do tipo pi-pi e pi-alkil, que, embora menos intensas que as interações cátion-pi, configuraram um arranjo espacial capaz de mimetizar de forma eficiente essas interações críticas estabelecidas pelo DNP com W86 e Y337. Esse alinhamento estrutural, associado à semelhança nos parâmetros eletrônicos, justificou o desempenho observado para PRT no sítio catalítico da hAChE. É importante ressaltar que, embora os resultados de docking molecular com conformações de ligantes calculadas em nível quântico forneçam resultados promissores, existem técnicas que complementam a análise de docking, proporcionando uma descrição mais completa do sistema biológico. Um exemplo é a abordagem que combina docking com perturbação da energia livre (do inglês *free-energy perturbation* — FEP), técnica esta que fornece a densidade de probabilidade dos estados ligados entre os complexos proteína-ligante, permitindo o cálculo de propriedades termodinâmicas de interesse

(Stock *et al.*, 2017). Outra aplicação relevante é o uso de simulações de dinâmica molecular, que em um estudo com o complexo DNP-hAChE, demonstraram múltiplos modos de ligação à temperatura ambiente (Silva *et al.*, 2020). Tal detalhamento sobre os mecanismos de inibição não seria capturado apenas por métodos de docking estático.

5.2 Data analytics

5.2.1 Principais PCs para o conjunto de dados estudado

Por meio da aplicação da PCA, foi possível identificar os descritores de maior relevância, considerando os dados ADMET e de estrutura eletrônica analisados, para o conjunto de dados utilizado. A partir da análise dos *loadings* e da variância explicada cumulativa pelas três primeiras componentes principais (PCs), foram selecionados os seguintes descritores: LUMO+1 (eV), Fração Csp³, número de aceptores de ligação de hidrogênio (Ligação-H^a) e Score (ADT), os quais, em conjunto, explicaram uma variância cumulativa de 94,21%. A Figura 5.17 ilustra a evolução da variância explicada cumulativa, comparando o conjunto inicial de dados com o conjunto final de dimensão reduzida, composto pelos descritores de maior peso.

Figura 5.17: A esquerda variância explicada cumulativa de 63,57% considerando todos os descritores e a direita variância explicada cumulativa de 94,21% após a redução da dimensionalidade dos dados.

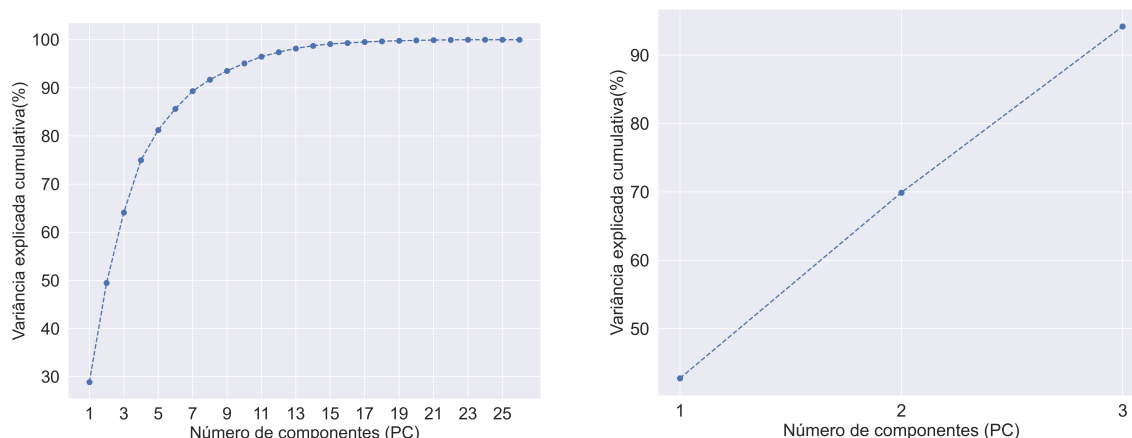
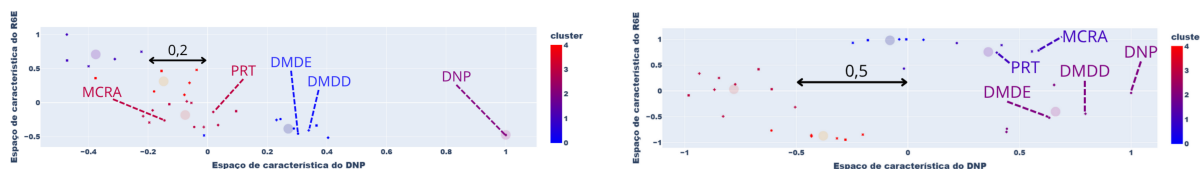


Figura 5.18: Espaço de características gerado pelo algoritmo K-Means antes (esquerda) e após (direita) a redução de dimensionalidade. A seleção dos descritores com maior variância explicada nas três primeiras componentes principais promoveu uma reorganização expressiva dos clusters, resultando na correta separação de CMG e LVD e na aproximação de moléculas ao *lead* DNP, que anteriormente se encontrava em um cluster isolado.



A Figura 5.18 ilustra a transformação do espaço de características gerado pelo algoritmo K-means após a redução da dimensionalidade. Inicialmente, quando todos os descritores foram considerados, o DNP permaneceu isolado em um único agrupamento, enquanto as moléculas com melhor perfil ADMET e estrutura eletrônica (como PRT, DMDE e DMDD) encontravam-se consideravelmente distantes no espaço de características. Após a redução da dimensionalidade para os quatro descritores mais relevantes, observou-se uma aproximação significativa dessas moléculas em relação ao DNP, acompanhada por um aumento na dispersão do espaço de características de 0,2 para 0,5 (distâncias entre clusters calculadas a partir da primeira molécula do cluster adjacente). Essa reorganização espacial evidencia a melhoria na qualidade da clusterização proporcionada pela seleção otimizada de descritores. As equações que

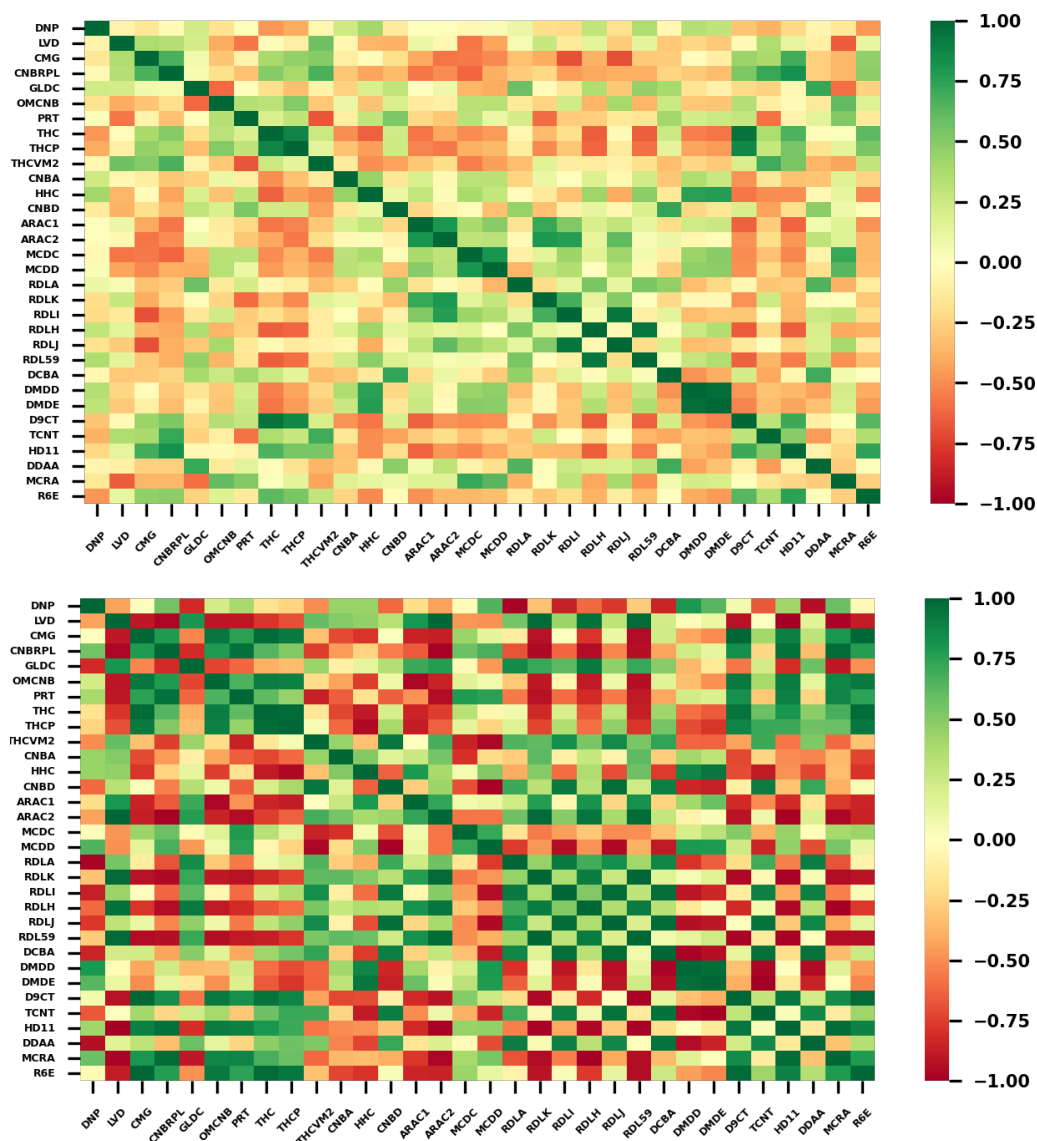
descrevem as três primeiras componentes principais para os descritores selecionados são:

$$\begin{aligned} PC_1 &= 0,72_{LUMO+1} + 0,67_{Csp^3} - 0,15_{Ligação-H^a} + 0,12_{Score(ADT)} \\ PC_2 &= 0,10_{LUMO+1} - 0,14_{Csp^3} + 0,52_{Ligação-H^a} + 0,84_{Score(ADT)} \\ PC_3 &= -0,07_{LUMO+1} + 0,33_{Csp^3} + 0,83_{Ligação-H^a} - 0,45_{Score(ADT)} \end{aligned} \quad (5.1)$$

Os resultados corroboram a importância do orbital de fronteira LUMO + 1 (eV) na mimetização das interações não covalentes do DNP, bem como destacam as contribuições das ligações de hidrogênio, da fração de carbonos com hibridização sp^3 (Csp^3) e da afinidade de interação — expressa pelo Score (ADT) — como os descritores que melhor explicam a variabilidade do conjunto de dados analisado.

Observou-se também que a seleção dos descritores de maior peso resultou no aumento das correlações, tanto positivas — como observado na classe das radulaninas — quanto negativas, a exemplo da relação entre LVD e CMG, que são moléculas estruturalmente distintas, mas que, inicialmente, considerando todos os descritores, apresentavam uma fraca correlação positiva. Esse ajuste nas matrizes de correlação (Figura 5.19) contribui para uma melhor definição dos agrupamentos, favorecendo a separação e a clusterização das classes no espaço de características, como no próprio composto *lead* DNP que inicialmente foi agrupado sozinho no cluster de cor roxa mas após a redução da dimensionalidade, foi agrupado no mesmo grupo do MCDD, MCRA, CNBRPL, PRT e HD11 (Figura 5.18).

Figura 5.19: Heatmaps com todos os descritores (superior) e com os descritores de maior peso após redução da dimensionalidade (inferior).



É importante ressaltar que, apesar dos resultados promissores, o método implementado para seleção dos melhores descritores ainda possui caráter empírico e qualitativo, sendo possível que a escolha de outros descritores conduza a resultados similares. Portanto, um dos nossos objetivos futuros consiste no desenvolvimento de um algoritmo baseado em heurísticas que formalize esta análise.

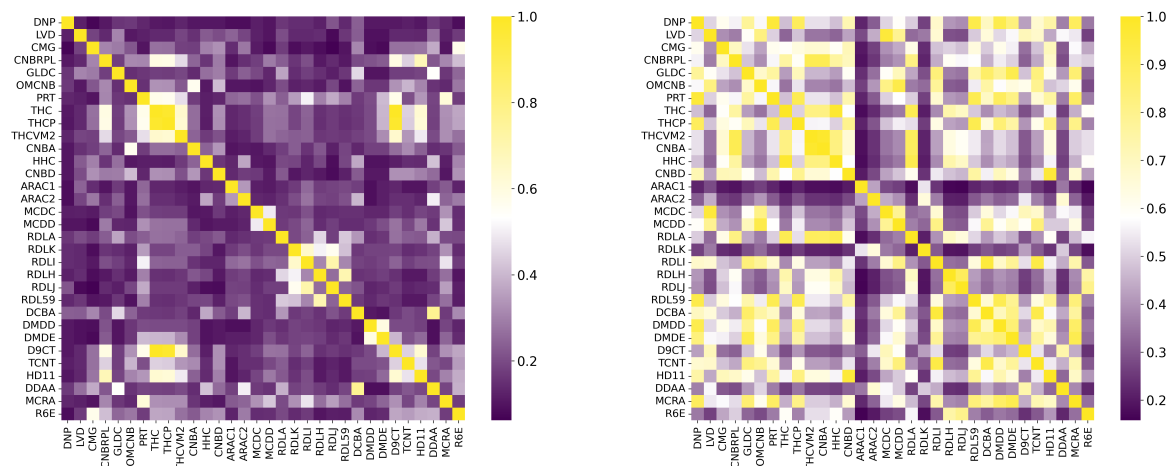
Para tanto, pretendemos implementar simulações automatizadas para a seleção otimizada de descritores, utilizando algoritmos heurísticos consolidados como os algoritmos genéticos (John, 2019) ou métodos de Monte Carlo (Metropolis *et al.*, 1953). O objetivo principal será realizar uma busca sistemática no espaço das componentes principais, visando identificar a combinação de descritores que maximize a variância explicada pelas três primeiras componentes principais.

5.3 Redes neurais artificiais

5.3.1 Rede neural siamesa-NCIVISION

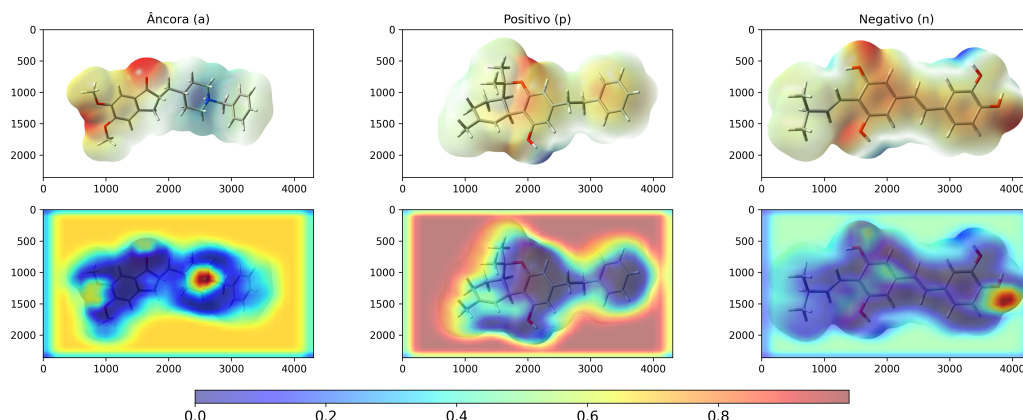
Os resultados obtidos a partir da rede siamesa com conjunto de dados de entrada de imagens de MEP e diagramas RDG, revelaram padrões de similaridade que não foram capturados pelos descritores eletrônicos globais e nem pela similaridade estrutural de Tanimoto S_T convencional (Figura 5.20). O composto PRT, que apresentou a maior afinidade de ligação no protocolo ADT ($-10,34$ kcal/mol), também foi identificado pelo modelo com a maior similaridade predita em relação ao DNP ($d^{-1} = 0.959$). Esse valor contrasta fortemente com a baixa similaridade de Tanimoto S_T de apenas 0,134, evidenciando que, embora estruturalmente distinto, o PRT compartilha com o DNP elevada similaridade em relação ao mapeamento do potencial eletrostático.

Figura 5.20: Similaridade de Tanimoto via SMILES à esquerda e similaridade predita pelo modelo a direita (via distância invertida, d^{-1})



Por outro lado, os compostos como o ARAC1, que apresentaram descritores eletrônicos bastante próximos aos do DNP— incluindo gap, HOMO/LUMO e índices de densidade condensada (CDD^+ e CDD^-) —, foram classificados pela rede com a menor similaridade ($d^{-1} = 0,220$). Este resultado explica seu fraco desempenho observado nos experimentos de docking (afinidade de $-8,09$ kcal/mol) em relação a faixa esperada do DNP, ainda que sem diferenças significativas nos descritores eletrônicos globais. A análise do Grad-CAM (Figura 5.21) evidenciou que, embora haja sobreposição geométrica entre o anel aromático ligado ao nitrogênio protonado do DNP e a região catecólica do ARAC1, o potencial eletrostático mais negativo da hidroxila no catecol de ARAC1 é interpretado pelo modelo como uma região desfavorável, comprometendo a formação de interações cátion-pi cruciais no sítio ativo da hAChE.

Figura 5.21: Grad-CAM com os resultados para DNP ($\hat{\text{Âncora-a}}$); PRT (Positivo-p) e ARAC1 (Negativo-n)



Já os compostos DCBA ($d^{-1} = 0,700$) e HD11 ($d^{-1} = 0,750$), apresentaram valores intermediários de similaridade, coerentes com suas baixas afinidades nos estudos de docking ($-7,14$ kcal/mol para DCBA e $-8,39$ kcal/mol para HD11). Esses compostos, embora tenham menor similaridade eletrostática do que o PRT, ainda compartilham regiões no mapa MEP favoráveis a interação com o sítio ativo da enzima. Os compostos THCVM2 ($d^{-1} = 0,528$) e R6E ($d^{-1} = 0,357$), obtiveram similaridades baixas, alinhando-se com seu baixo desempenho nos estudos de docking ($-4,99$ kcal/mol para o THCVM2 e $-7,15$ kcal/mol para o R6E). Este resultado reforça que, apesar de possuírem gaps e descritores eletrônicos comparáveis aos do DNP, a distribuição espacial do potencial eletrostático nestas moléculas não favorece a interação eficaz com a hAChE.

A análise combinada da distância invertida predita pelo modelo, das afinidades de interação de docking e dos descritores de estrutura eletrônica, demonstrou que o modelo foi capaz de capturar características locais e regionais dos MEP, determinantes para a afinidade de ligação. Em especial, o contraste observado pelo PRT, com máxima similaridade, e o ARAC1, com mínima similaridade, exemplifica como o modelo é capaz de diferenciar não apenas a geometria, mas sobretudo as nuances do ambiente eletrostático de cada molécula, que modulam as interações moleculares no sítio ativo da enzima, utilizando além dos MEP, os *embeddings* de 64 dimensões gerados com base nas matrizes GLCM obtidas através dos diagramas RDG (Figura 5.23).

Estes resultados indicam que a utilização de modelos de redes neurais baseados em MEP é uma estratégia complementar robusta para análises tradicionais de docking e de propriedades eletrônicas, sendo particularmente eficaz na identificação de padrões de similaridade funcional que não podem ser explicados por métodos convencionais baseados apenas na estrutura química, ou que analisam os resultados de docking e de estrutura eletrônica de forma isolada. É importante destacar que, por se tratar de uma metodologia inédita, torna-se necessária a construção de bases de dados com MEP e diagramas RDG, visando ampliar o conjunto de dados e aprimorar a acurácia do modelo.

Figura 5.22: 32 diagramas RDG de entrada para o modelo.

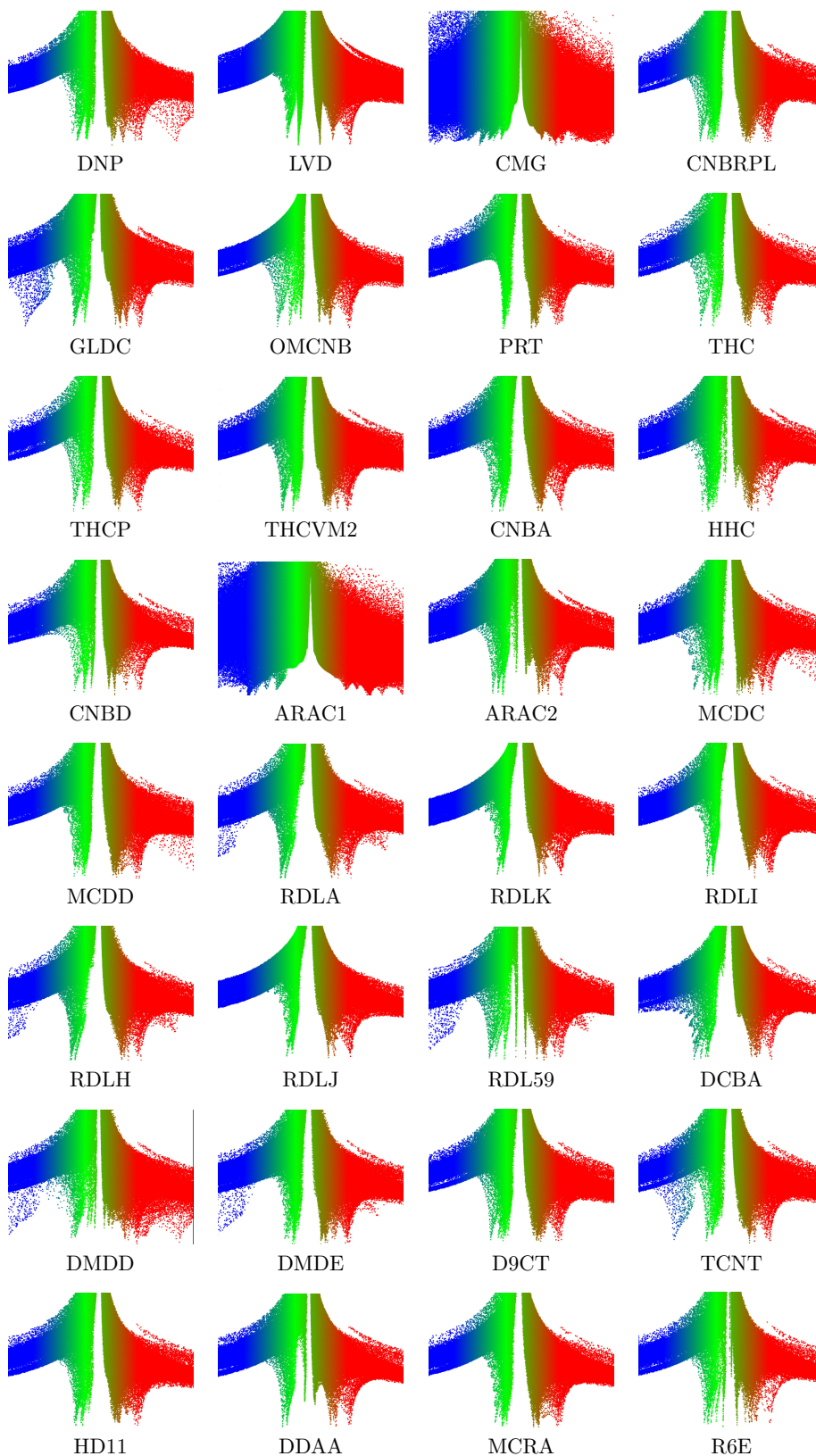


Figura 5.23: Mosaico dos *embeddings* com 64 dimensões gerados pelo modelo a partir dos diagramas RDG para os 32 compostos analisados no estudo.

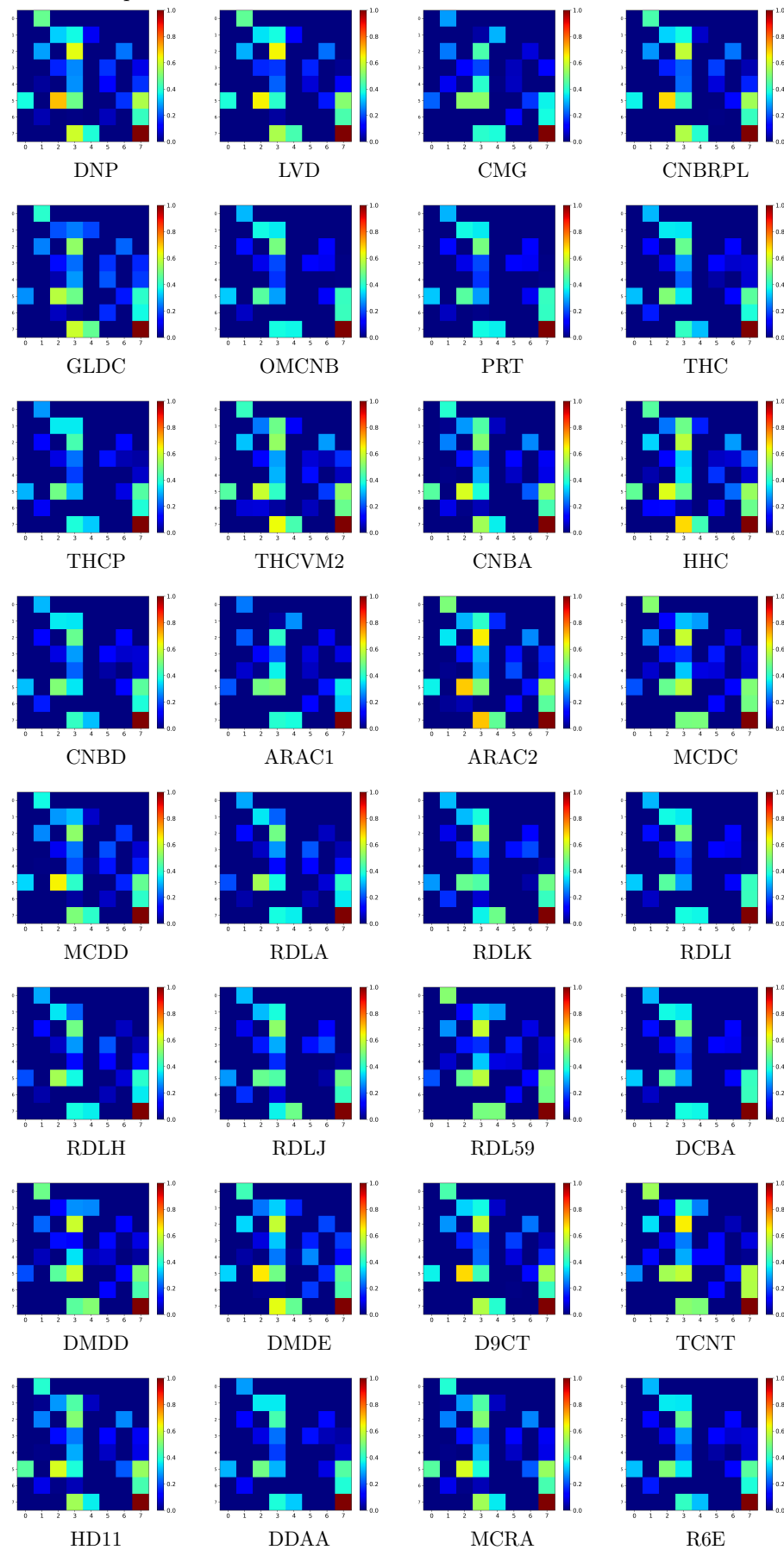
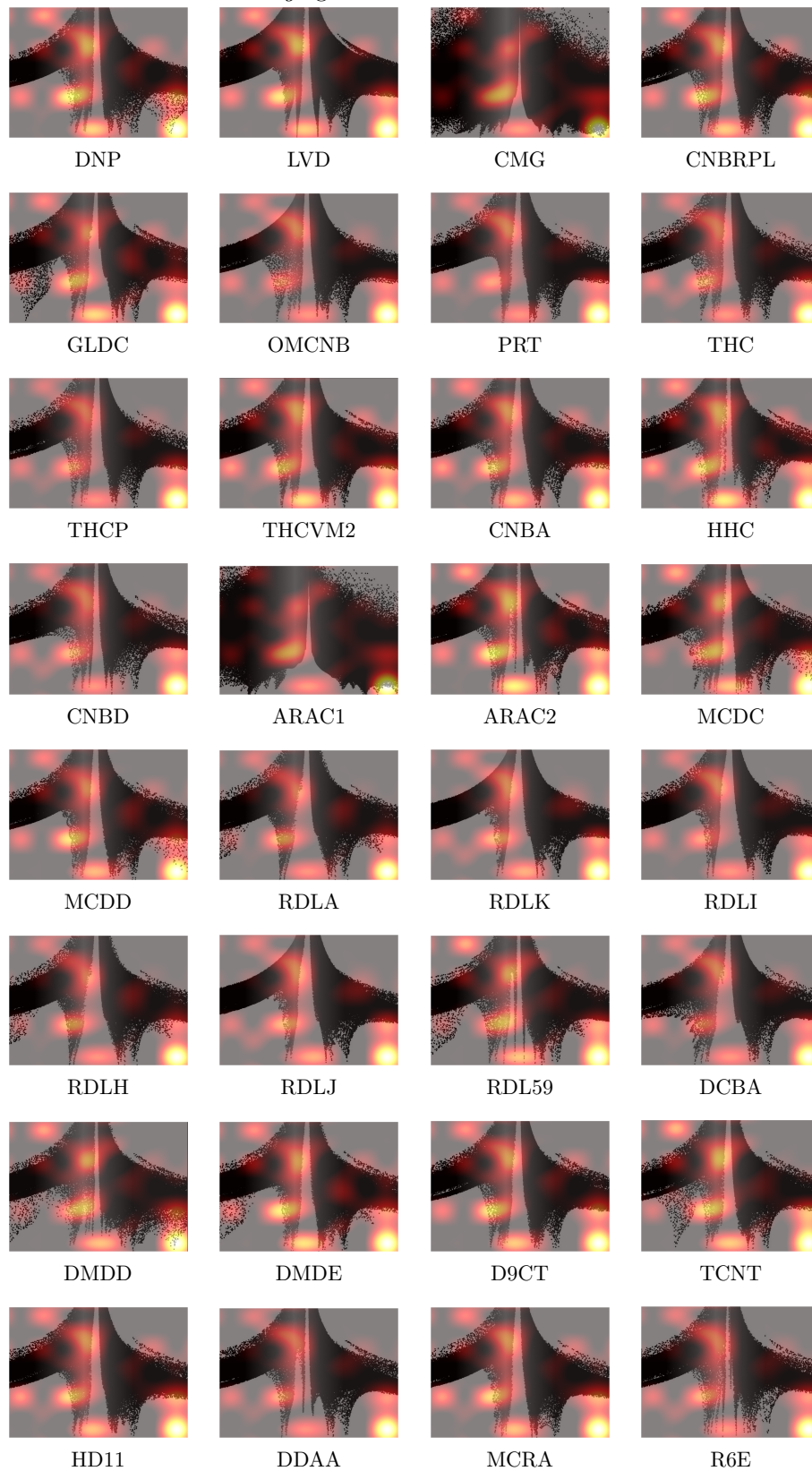


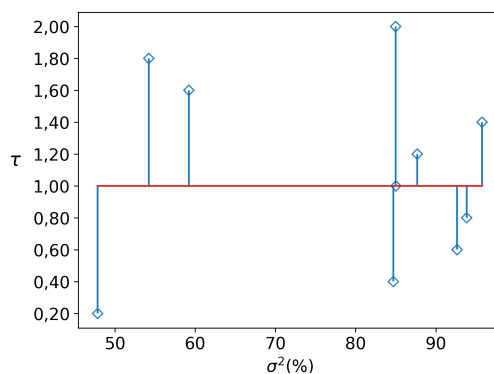
Figura 5.24: Visualização das regiões dos diagramas RDG que receberam maior atenção do modelo, de acordo com as matrizes de *embeddings* geradas.



5.3.2 Autoencodificador variacional generativo acoplado a células de memória de curta e longa duração-VAELSTM

Os espaços latentes gerados para cada temperatura, após a redução da dimensionalidade para as duas primeiras componentes principais, demonstram a influência da temperatura no processo generativo. A variância total explicada por essas duas componentes não apresentou uma tendência linear de crescimento com o hiperparâmetro temperatura, τ , havendo picos em temperaturas intermediárias como $\tau = 0,8$ com $\sigma^2 = 93,85\%$ e $\tau = 1,4$ com $\sigma^2 = 95,77\%$ (Figura 5.25). Esses resultados sugerem que, em certas faixas de temperatura, as duas primeiras componentes principais, PC1 e PC2, foram particularmente relevantes na informação retida pelo modelo no espaço latente. Por outro lado, em temperaturas muito baixas, como $\tau = 0,2$ com $\sigma^2 = 47,82\%$, ou ligeiramente mais elevadas, como $\tau = 1,6$ com $\sigma^2 = 59,21\%$ e $\tau = 1,8$ com $\sigma^2 = 54,21\%$, a variância explicada foi consideravelmente menor, indicando uma menor diversidade molecular capturada por essas componentes.

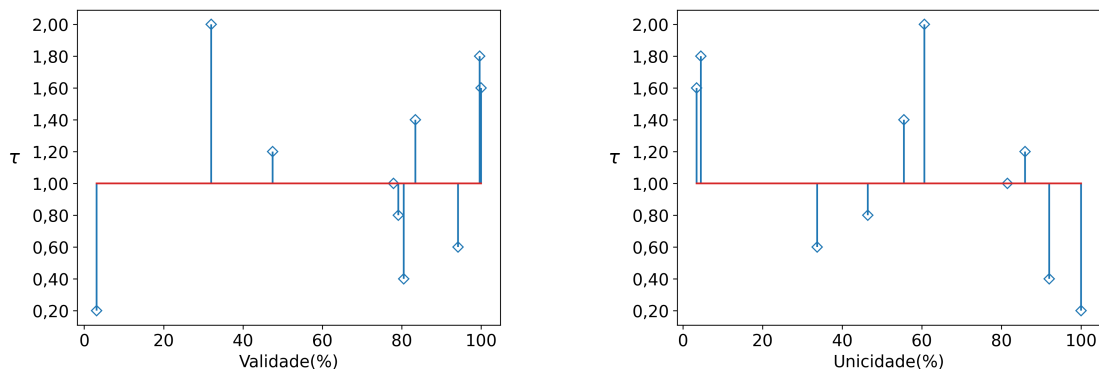
Figura 5.25: Gráfico de Hastes mostrando a relação entre o hiperparâmetro τ e a variância.



Essas flutuações e a não linearidade na relação entre a temperatura e a variância explicada são aspectos amplamente discutidos na literatura (Holtzman *et al.*, 2019). A complexidade da topologia do espaço latente gerado, aliada à possibilidade de modos colapsados locais ou à geração de um número elevado de moléculas inválidas em função de uma amostragem excessivamente “criativa” em temperaturas altas, pode levar a essas variações (Hu *et al.*, 2023b; Arjovsky *et al.*, 2017).

Os resultados de validade e unicidade das moléculas geradas ao longo das diferentes temperaturas reforçam a complexidade da definição do hiperparâmetro τ para garantir a qualidade do processo generativo (Figura 5.26). Em temperaturas muito baixas, como $\tau = 0,2$, a alta unicidade (100%) contrastou com a baixa validade (3,1%), indicando que, embora cada molécula válida seja única, o modelo gera moléculas válidas com baixa frequência — possivelmente em razão da amostragem excessivamente restrita decorrente da baixa temperatura. Por outro lado, em temperaturas intermediárias, como $\tau = 0,4$, $\tau = 1,0$ e $\tau = 1,4$, observou-se um equilíbrio mais desejável entre validade (80,5%, 77,9% e 83,4%, respectivamente) e unicidade (92%, 81,5% e 55,8%, respectivamente).

Figura 5.26: Gráfico de hastes mostrando a relação entre o hiperparâmetro τ com os parâmetros Validade e Unicidade.



Temperaturas intermediárias como $\tau = 0,6$, $\tau = 0,8$ e $\tau = 1,4$ se destacaram por apresentarem alta validade (acima de 79%), níveis razoáveis de unicidade (entre 33% e 55%) e, sobretudo, os maiores percentuais de variância explicada pelas duas primeiras componentes principais (92,65%, 93,85% e 95,77%, respectivamente). Esses resultados indicam que, nessas faixas de temperatura, o modelo não apenas gera moléculas válidas e relativamente diversas, como também estrutura o espaço latente de maneira mais interpretável e concentrada em duas dimensões principais — o que facilita análises visuais e semânticas. É fundamental destacar que o modelo proposto, em seu estágio atual de desenvolvimento, apresenta características parcialmente semânticas, demonstrando capacidade de capturar relações sequenciais específicas na representação SMILES durante o processo generativo. Contudo, para alcançar um modelo completamente semântico — capaz de compreender plenamente o significado dos dados, particularmente em contextos linguísticos — seria necessária a implementação de mecanismos de atenção mais sofisticados. Essa evolução transcenderia as limitações inerentes à arquitetura de memória de curta e longa duração das redes LSTM, aproximando-se das capacidades observadas na arquitetura Transformer (Vaswani *et al.*, 2017), amplamente utilizada em *large language models* (LLM). Esforços nesta direção já vêm sendo empregados pela comunidade científica (Sadeghi *et al.*, 2024).

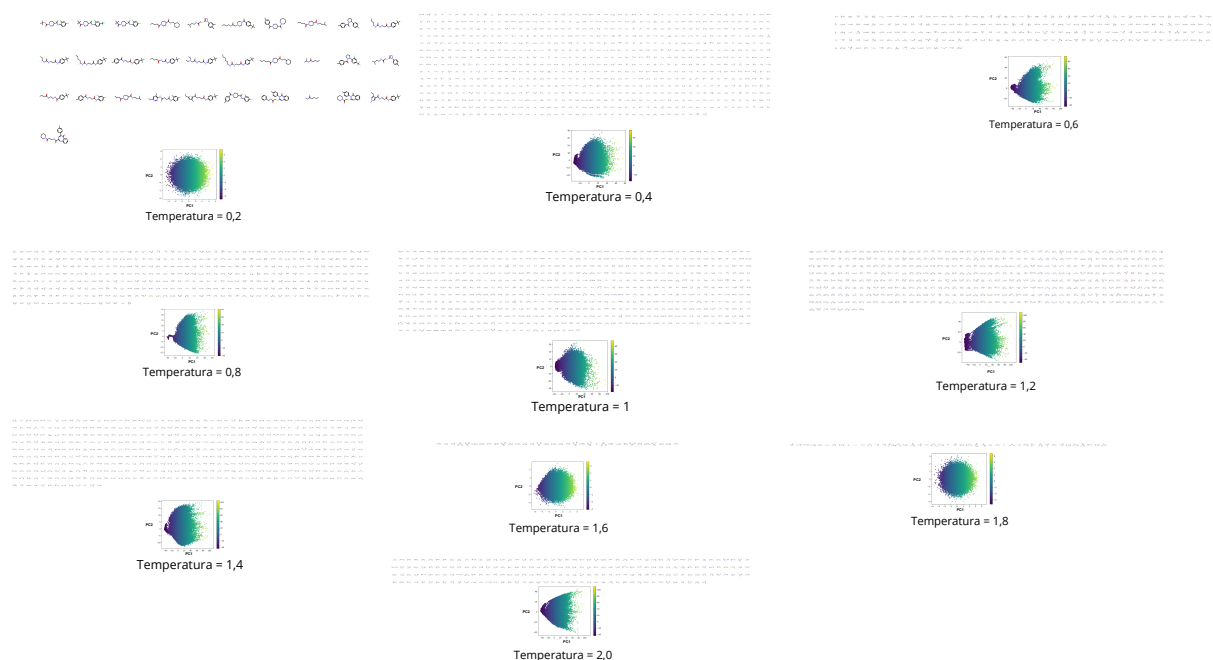
Em contrapartida, temperaturas mais elevadas como $\tau = 1,6$ e $\tau = 1,8$, apesar de alcançarem validade próxima ou superior a 99%, apresentaram baixa unicidade (3,4% e 4,5%, respectivamente) e baixa variância explicada pelas PCs (59,21% e 54,21%). Isso evidencia a tendência do modelo a colapsar em soluções redundantes, com menor diversidade estrutural e maior complexidade no espaço latente. Assim, conclui-se que a faixa ótima de temperatura para o experimento realizado situou-se entre aproximadamente 0,6 e 1,4 — intervalo no qual o modelo, de modo geral, balanceou bem a validade, a unicidade e a organização do espaço latente conforme indicado pela variância explicada pelas duas primeiras componentes principais.

Em relação à distribuição do espaço latente, observa-se na Figura 5.27 que a temperatura teve um impacto expressivo em sua forma e dispersão. Nas temperaturas $\tau = 0,2$, $\tau = 1,6$ e $\tau = 1,8$, a distribuição dos pontos foi visivelmente mais compacta e simétrica, com os eixos das duas primeiras componentes principais operando em faixas de valores de $\log P$ limitadas. O espaço latente assumiu um formato mais arredondado ou elíptico, sugerindo que o modelo, nessas condições de amostragem, explorou uma região mais restrita do espaço de probabilidade — o que se refletiu em menor diversidade molecular (valores de unicidade para $\tau = 1,6$ e $\tau = 1,8$ iguais a 3,38% e 4,52%, respectivamente; e $\tau = 0,2$, apesar de ter apresentado 100% de unicidade, teve validade de apenas 3,1%).

A relação entre o $\log P$ e as dimensões latentes foi consistente ao longo de todas as temperaturas, sendo que a primeira componente, de modo geral, apresentou maior variabilidade na faixa de valores de $\log P$, como se observa nos diagramas de dispersão da Figura 5.27. Outro aspecto relevante é que a dispersão dos valores de $\log P$ acompanhou a expansão do espaço latente: para valores de τ como 0,2 e 1,6, onde o espaço gerado foi mais compacto, o gradiente de $\log P$ também foi mais restrito, variando aproximadamente entre -3 e 3. Em contraste, para temperaturas intermediárias e mais altas, o espaço latente assumiu uma forma característica, semelhante a um “cometa”, com uma região mais densa e concentrada — a “cabeça” — contendo valores menores de $\log P$, possivelmente representando aglomerados de compostos moleculares semelhantes ou de baixa diversidade. A partir dessa região, observa-se uma “cauda” mais dispersa, com gradiente gradual para o verde e amarelo, indicando maiores valores de $\log P$. Essa continuidade é uma característica desejável em espaços latentes, pois favorece processos de interpolação e otimização.

Ademais, não foram observados grandes vazios no espaço latente, o que é um bom indicativo de que o modelo está explorando regiões relevantes para a geração molecular.

Figura 5.27: Moléculas únicas geradas para cada temperatura, τ , para o experimento inicial com conjunto de 310253 SMILES.



Capítulo 6

Conclusões

Este trabalho teve como objetivo realizar estudos de docking molecular, com ênfase em um conjunto de 32 moléculas classificadas como canabinóides, canabinóides raros e canabinomiméticos, bem como aplicar e implementar algoritmos de machine learning no processo de descoberta de novos inibidores da enzima acetilcolinesterase humana (hAChE),

O estudo de docking revelou um padrão na mimetização das interações do tipo cátion-pi entre o ligante de referência, DNP, e os resíduos do sítio ativo da hAChE W86 e Y337, principalmente na região do ligante onde se encontra distribuído o orbital de fronteira LUMO+1, esse padrão foi observado nas moléculas avaliadas com valores de score maiores que -7,5 kcal/mol. As principais interações observadas entre os complexos formados entre os resíduos do GORGE e os compostos foram interações do tipo pi, tais como pi-alquil, pi-pi T-shaped, pi-pi stacked e pi-sigma. Dentre as 32 moléculas estudadas, sete se destacaram por apresentarem os melhores valores de energia de interação e maior similaridade conformacional em relação ao DNP. Entre elas, ressaltou-se a molécula perrotetineno (PRT), que mimetizou de forma mais eficaz as interações cátion-pi e obteve o valor de score mais próximo ao DNP.

A análise de componentes principais demonstrou que os descritores LUMO+1, Csp^3 , número de aceptores de ligação de hidrogênio e o valor da energia de interação dos complexos formados foram os mais relevantes para a variância total do conjunto, o que foi corroborado pelos resultados da matriz de similaridade e do k-means. Tais métodos revelaram maior coerência no agrupamento molecular, destacando a correta separação das entidades moleculares do levodopa (LVD) e do carmagerol (CMG) como compostos dissimilares. Além disso, observou-se uma melhora na clusterização no espaço de características, com aumento de integrantes no grupo do DNP e separação clara entre CMG e LVD.

O modelo de visão computacional NCIVISION, voltado à determinação de similaridade, foi capaz de atribuir valores compatíveis com os obtidos nas análises de dados, especialmente nos casos dos compostos PRT e ARAC1. O modelo mostrou-se capaz de inferir similaridade a partir das imagens dos MEP e dos diagramas RDG, associando propriedades de distribuição eletrostática e interações não covalentes. Contudo, destaca-se a necessidade de estudos mais aprofundados com um conjunto de dados ampliado, a fim de aumentar a robustez e a validação do modelo.

O modelo generativo mostrou-se eficiente na geração de moléculas válidas a partir de conjuntos de dados de SMILES diversos, controlando a diversidade por meio do hiperparâmetro de temperatura, τ . No entanto, torna-se necessário avaliar com maior profundidade as propriedades AChEIs das moléculas geradas.

Referências

- ABD EL-KARIM, S. S.; ANWAR, M. M.; AHMED, N. S.; SYAM, Y. M.; ELSEGINY, S. A.; ALY, H. F.; YOUNIS, E. A.; KHALIL, W. K.; AHMED, K. A.; MOHAMMED, F. F. *et al.* Discovery of novel benzofuran-based derivatives as acetylcholinesterase inhibitors for the treatment of Alzheimer's disease: Design, synthesis, biological evaluation, molecular docking and 3D-QSAR investigation. **European Journal of Medicinal Chemistry**, Elsevier, v. 260, p. 115766, 2023. DOI: <https://doi.org/10.1016/j.ejmech.2023.115766>. Citado na p. 25.
- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley interdisciplinary reviews: computational statistics**, Wiley Online Library, v. 2, n. 4, p. 433–459, 2010. Citado na p. 43.
- ABIWINANDA, N.; HANIF, M.; HESAPUTRA, S. T.; HANDAYANI, A.; MENGKO, T. R. Brain tumor classification using convolutional neural network. *In: SPRINGER. WORLD Congress on Medical Physics and Biomedical Engineering 2018: June 3-8, 2018, Prague, Czech Republic (Vol. 1). [S. l.: s. n.]*, 2019. p. 183–189. Citado na p. 47.
- ABREU SILVA, M. de; SETTE, C. D. B.; KIAMETIS, A. S.; ROMEIRO, L. A. S.; GARGANO, R. Molecular modeling of cardanol-derived AChE inhibitors. **Chemical Physics Letters**, Elsevier, v. 731, p. 136591, 2019. DOI: <https://doi.org/10.1016/j.cplett.2019.07.019>. Citado na p. 18.
- ACKLEY, D. H.; HINTON, G. E.; SEJNOWSKI, T. J. A learning algorithm for Boltzmann machines. **Cognitive science**, Elsevier, v. 9, n. 1, p. 147–169, 1985. Citado na p. 44.
- ADAMS, H. K.; KADARAUCH, M.; HODSON, N. J.; LIT, A. R.; PHIPPS, R. J. Design Approaches That Utilize Ionic Interactions to Control Selectivity in Transition Metal Catalysis. **Chemical Reviews**, ACS Publications, v. 125, n. 5, p. 2846–2907, 2025. DOI: <https://doi.org/10.1021/acs.chemrev.4c00849>. Citado na p. 29.
- ADLIMOGHADDAM, A.; NEUENDORFF, M.; ROY, B.; ALBENSI, B. C. A review of clinical treatment considerations of donepezil in severe Alzheimer's disease. **CNS neuroscience & therapeutics**, Wiley Online Library, v. 24, n. 10, p. 876–888, 2018. Citado na p. 17.
- AHMED, T. F.; AHMED, A.; IMTIAZ, F. History in perspective: How Alzheimer's Disease came to be where it is? **Brain Research**, Elsevier, v. 1758, p. 147342, 2021. Citado na p. 13.
- ALAVI NAINI, S. M.; SOUSSI-YANICOSTAS, N. Tau hyperphosphorylation and oxidative stress, a critical vicious circle in neurodegenerative tauopathies? **Oxidative medicine and cellular longevity**, Wiley Online Library, v. 2015, n. 1, p. 151979, 2015. DOI: <https://doi.org/10.1155/2015/151979>. Citado na p. 14.
- ALCOLEA-PALAFOX, M.; POSADA-MORENO, P.; ORTUÑO-SORIANO, I.; PACHECO-DEL-CERRO, J. L.; MARTÍNEZ-RINCÓN, C.; RODRÍGUEZ-MARTÍNEZ, D.; PACHECO-CUEVAS, L. Research strategies developed for the treatment of alzheimer's disease. Reversible and pseudo-irreversible inhibitors of acetylcholinesterase: Structure-activity relationships and drug design. *In: DRUG design and discovery in Alzheimer's disease. [S. l.]: Elsevier, 2014. p. 426–477. Citado na p. 17.*

- ALONSO, D.; DORRONSORO, I.; RUBIO, L.; MUNOZ, P.; GARCÍA-PALOMERO, E.; DEL MONTE, M.; BIDON-CHANAL, A.; OROZCO, M.; LUQUE, F.; CASTRO, A. *et al.* Donepezil–tacrine hybrid related derivatives as new dual binding site inhibitors of AChE. **Bioorganic & medicinal chemistry**, Elsevier, v. 13, n. 24, p. 6588–6597, 2005. Citado na p. 19.
- ALZHEIMER'S DISEASE, S. of. **Stages of Alzheimer's Disease**. [*S. l.: s. n.*], 2019. <https://alzheimersdisease.net/living/stages>. Accessed: 2024-05-16. Citado na p. 13.
- AMEN, Y. M.; MARZOUK, A. M.; ZAGHLOUL, M. G.; AFIFI, M. S. The genus *Machaerium* (Fabaceae): taxonomy, phytochemistry, traditional uses and biological activities. **Natural product research**, Taylor & Francis, v. 29, n. 15, p. 1388–1405, 2015. DOI: <https://doi.org/10.1080/14786419.2014.1003062>. Citado na p. 24.
- ANDRADE RAMOS, G. de; OLIVEIRA, A. S. de; BARTOLINI, M.; NALDI, M.; LIPARULO, I.; BERGAMINI, C.; ULIASSI, E.; WU, L.; FRASER, P. E.; ABREU, M. *et al.* Discovery of sustainable drugs for Alzheimer's disease: cardanol-derived cholinesterase inhibitors with antioxidant and anti-amyloid properties. **RSC Medicinal Chemistry**, Royal Society of Chemistry, v. 12, n. 7, p. 1154–1163, 2021. DOI: <https://doi.org/10.1039/d1md00046b>. Citado na p. 18.
- ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein generative adversarial networks. *In*: PMLR. INTERNATIONAL conference on machine learning. [*S. l.: s. n.*], 2017. p. 214–223. Citado na p. 72.
- ARNOLD, S. E.; ARVANITAKIS, Z.; MACAULEY-RAMBACH, S. L.; KOENIG, A. M.; WANG, H.-Y.; AHIMA, R. S.; CRAFT, S.; GANDY, S.; BUETTNER, C.; STOECKEL, L. E. *et al.* Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. **Nature Reviews Neurology**, Nature Publishing Group UK London, v. 14, n. 3, p. 168–181, 2018. DOI: <https://doi.org/10.1038/nrneuro.2017.185>. Citado na p. 15.
- ASAKAWA, Y. Bryophytes Their Chemistry and Chemotaxonomy of Bryophytes. (**No Title**), p. 369, 1990. Citado na p. 22.
- ASAKAWA, Y.; NAGASHIMA, F.; LUDWICZUK, A. Distribution of bibenzyls, prenyl bibenzyls, bis-bibenzyls, and terpenoids in the liverwort genus *Radula*. **Journal of Natural Products**, ACS Publications, v. 83, n. 3, p. 756–769, 2020. Citado na p. 22.
- ASAKAWA, Y.; TAKIKAWA, K.; TOYOTA, M.; TAKEMOTO, T. Novel bibenzyl derivatives and ent-cuparene-type sesquiterpenoids from *Radula* species. **Phytochemistry**, Elsevier, v. 21, n. 10, p. 2481–2490, 1982. Citado na p. 23.
- ATKINS, P. W.; DE PAULA, J.; KEELER, J. **Atkins' physical chemistry**. [*S. l.*]: Oxford university press, 2023. Citado na p. 28.
- BECKE, A. D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. **The Journal of chemical physics**, American Institute of Physics, v. 96, n. 3, p. 2155–2160, 1992. Citado na p. 49.
- BHADWAL, A. S.; KUMAR, K.; KUMAR, N. NRC-VABS: Normalized Reparameterized Conditional Variational Autoencoder with applied beam search in latent space for drug molecule design. **Expert Systems with Applications**, Elsevier, v. 240, p. 122396, 2024. Citado na p. 46.
- BHUNIA, S.; KOLISHETTI, N.; ARIAS, A. Y.; VASHIST, A.; NAIR, M. Cannabidiol for neurodegenerative disorders: A comprehensive review. **Frontiers in pharmacology**, Frontiers Media SA, v. 13, p. 989717, 2022. DOI: <https://doi.org/10.3389/fphar.2022.989717>. Citado na p. 22.
- BIRKS, J. S.; EVANS, J. G. Rivastigmine for Alzheimer's disease. **Cochrane Database of systematic reviews**, John Wiley & Sons, Ltd, n. 4, 2015. DOI: <https://doi.org/10.1002/14651858.cd001191.pub3>. Citado na p. 17.

BISHOP, C. M.; NASRABADI, N. M. **Pattern recognition and machine learning**. [S. l.]: Springer, 2006. v. 4. Citado nas pp. 38, 40, 43.

BLUMENFELD, J.; YIP, O.; KIM, M. J.; HUANG, Y. Cell type-specific roles of APOE4 in Alzheimer disease. **Nature Reviews Neuroscience**, Nature Publishing Group UK London, v. 25, n. 2, p. 91–110, 2024. DOI: <https://doi.org/10.1038/s41583-023-00776-9>. Citado na p. 15.

BONCOEUR, É.; BOUVET, G. F.; MIGNEAULT, F.; TARDIF, V.; FERRARO, P.; RADZIOCH, D.; SANCTIS, J. B. de; EIDELMAN, D.; GOVINDARAJU, K.; DAGENAIS, A. *et al.* Induction of nitric oxide synthase expression by lipopolysaccharide is mediated by calcium-dependent PKC α - β 1 in alveolar epithelial cells. **American Journal of Physiology-Lung Cellular and Molecular Physiology**, American Physiological Society Bethesda, MD, v. 305, n. 2, p. 1175–1184, 2013. Citado na p. 23.

BRADLEY, R. E.; D'ANTONIO, L. A.; SANDIFER, C. E. Euler at 300: an appreciation. (No Title), 2007. Citado na p. 44.

BRENEMAN, C. M.; WIBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. **Journal of computational chemistry**, Wiley Online Library, v. 11, n. 3, p. 361–373, 1990. Citado na p. 48.

BRIDLE, J. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. **Advances in neural information processing systems**, v. 2, 1989. Citado na p. 44.

BRUERA, E.; EL OSTA, B.; VALERO, V.; DRIVER, L. C.; PEI, B.-L.; SHEN, L.; POULTER, V. A.; PALMER, J. L. Donepezil for cancer fatigue: a double-blind, randomized, placebo-controlled trial. **Journal of clinical oncology**, American Society of Clinical Oncology, v. 25, n. 23, p. 3475–3481, 2007. Citado na p. 17.

BUCKINGHAM, A. D. Permanent and induced molecular moments and long-range intermolecular forces. **Advances in chemical physics: Intermolecular forces**, Wiley Online Library, p. 107–142, 1967. DOI: <https://doi.org/10.1002/9780470143582.ch2>. Citado na p. 29.

CABRERA-PARDO, J. R.; FUENTEALBA, J.; GAVILÁN, J.; CAJAS, D.; BECERRA, J.; NAPIÓRKOWSKA, M. Exploring the multi-target neuroprotective chemical space of benzofuran scaffolds: A new strategy in drug development for alzheimer's disease. **Frontiers in Pharmacology**, Frontiers Media SA, v. 10, p. 1679, 2020. DOI: <https://doi.org/10.3389/fphar.2019.01679>. Citado na p. 25.

CAO, C.; LI, Y.; LIU, H.; BAI, G.; MAYL, J.; LIN, X.; SUTHERLAND, K.; NABAR, N.; CAI, J. The potential therapeutic effects of THC on Alzheimer's disease. **Journal of Alzheimer's Disease**, SAGE Publications Sage UK: London, England, v. 42, n. 3, p. 973–984, 2014. DOI: <https://doi.org/10.3233/jad-140093>. Citado na p. 22.

CATAKLI, D.; ERZURUMLU, Y.; ASCI, H.; SAVRAN, M.; SEZER, S. Evaluation of cytoprotective effects of cannabidiol on neuroinflammation and neurogenesis process in rat offsprings. **Reproductive Toxicology**, Elsevier, v. 132, p. 108761, 2025. DOI: <https://doi.org/10.1016/j.reprotox.2024.108761>. Citado na p. 22.

CAZZATO, D.; CIMARELLI, C.; SANCHEZ-LOPEZ, J. L.; VOOS, H.; LEO, M. A survey of computer vision methods for 2d object detection from unmanned aerial vehicles. **Journal of Imaging**, MDPI, v. 6, n. 8, p. 78, 2020. Citado na p. 39.

CHANDLER, D. Interfaces and the driving force of hydrophobic assembly. **Nature**, Nature Publishing Group UK London, v. 437, n. 7059, p. 640–647, 2005. Citado na p. 30.

- CHANG, R.-F.; WU, W.-J.; MOON, W. K.; CHEN, D.-R. Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. **Ultrasound in medicine & biology**, Elsevier, v. 29, n. 5, p. 679–686, 2003. Citado na p. 32.
- CHEN, J.-Y.; ZHU, Q.; ZHANG, S.; OUYANG, D.; LU, J.-H. Resveratrol in experimental Alzheimer's disease models: A systematic review of preclinical studies. **Pharmacological Research**, Elsevier, v. 150, p. 104476, 2019. Citado na p. 26.
- CHEN, L.-G.; ZHANG, Y.-Q.; WU, Z.-Z.; HSIEH, C.-W.; CHU, C.-S.; WUNG, B.-S. Peanut arachidin-1 enhances Nrf2-mediated protective mechanisms against TNF- α -induced ICAM-1 expression and NF- κ B activation in endothelial cells. **International Journal of Molecular Medicine**, Spandidos Publications, v. 41, n. 1, p. 541–547, 2018. Citado na p. 26.
- CHEN, Z.-R.; HUANG, J.-B.; YANG, S.-L.; HONG, F.-F. Role of cholinergic signaling in Alzheimer's disease. **Molecules**, MDPI, v. 27, n. 6, p. 1816, 2022. Citado na p. 15.
- CHICCA, A.; SCHAFROTH, M. A.; REYNOSO-MORENO, I.; ERNI, R.; PETRUCCI, V.; CARREIRA, E. M.; GERTSCH, J. Uncovering the psychoactivity of a cannabinoid from liverworts associated with a legal high. **Science advances**, American Association for the Advancement of Science, v. 4, n. 10, eaat2166, 2018. Citado nas pp. 22, 23.
- CHON, A.; BALACHANDAR, N.; LU, P. Deep convolutional neural networks for lung cancer detection. **Stanford University**, p. 1–9, 2017. Citado na p. 47.
- CHONG, F. P.; NG, K. Y.; KOH, R. Y.; CHYE, S. M. Tau proteins and tauopathies in Alzheimer's disease. **Cellular and molecular neurobiology**, Springer, v. 38, p. 965–980, 2018. DOI: <https://doi.org/10.1007/s10571-017-0574-1>. Citado na p. 14.
- CHOUDHURY, M. G.; KUMARI, S.; DAS, K. B.; SAHA, N. Lipopolysaccharide causes NF κ B-mediated induction of inducible nitric oxide synthase gene and more production of nitric oxide in air-breathing catfish, *Clarias magur* (Hamilton). **Gene**, Elsevier, v. 658, p. 18–27, 2018. Citado na p. 23.
- CLAVAGUERA, F.; BOLMONT, T.; CROWTHER, R. A.; ABRAMOWSKI, D.; FRANK, S.; PROBST, A.; FRASER, G.; STALDER, A. K.; BEIBEL, M.; STAUFENBIEL, M. *et al.* Transmission and spreading of tauopathy in transgenic mouse brain. **Nature cell biology**, Nature Publishing Group UK London, v. 11, n. 7, p. 909–913, 2009. DOI: <https://doi.org/10.1038/ncb1901>. Citado na p. 14.
- COELHO, F.; BIRKS, J. Physostigmine for Alzheimer's disease. **The Cochrane database of systematic reviews**, n. 2, p. cd001499–cd001499, 2001. DOI: <https://doi.org/10.1002/14651858.cd001191.pub3>. Citado na p. 18.
- COHEN, K.; WEIZMAN, A.; WEINSTEIN, A. Modulatory effects of cannabinoids on brain neurotransmission. **European Journal of Neuroscience**, Wiley Online Library, v. 50, n. 3, p. 2322–2345, 2019. DOI: <https://doi.org/10.1111/ejn.14407>. Citado na p. 22.
- COLES, M.; STEINER-LIM, G. Z.; KARL, T. Therapeutic properties of multi-cannabinoid treatment strategies for Alzheimer's disease. **Frontiers in Neuroscience**, Frontiers Media SA, v. 16, p. 962922, 2022. Citado na p. 21.
- CRISMON, M. L. Tacrine: first drug approved for Alzheimer's disease. **Annals of Pharmacotherapy**, SAGE Publications, v. 28, n. 6, p. 744–751, 1994. DOI: <https://doi.org/10.1177/106002809402800612>. Citado na p. 17.
- CUEVA-CHAMBA, A.; BUSTAMANTE-PACHECO, F.; VANEGAS, D.; PEÑAHERRERA, E. Traditional medicinal uses and biological activities of species of the genus *Desmodium*: a literature review. **Boletín Latinoamericano y del Caribe de Plantas Medicinales y Aromáticas**, v. 22, n. 6, 2023. Citado na p. 25.

CULLMANN, F.; BECKER, H. Prenylated bibenzyls from the liverwort *Radula laxiramea*. **Zeitschrift für Naturforschung C**, Verlag der Zeitschrift für Naturforschung, v. 54, n. 3-4, p. 147–150, 1999. Citado na p. 23.

DAINA, A.; MICHELIN, O.; ZOETE, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. **Scientific reports**, Nature Publishing Group UK London, v. 7, n. 1, p. 42717, 2017. Citado na p. 49.

DALE, H. H. The action of certain esters and ethers of choline, and their relation to muscarine. **Journal of Pharmacology and Experimental Therapeutics**, ASPET, v. 6, n. 2, p. 147–190, 1914. Citado na p. 15.

DAULATZAI, M. A. Cerebral hypoperfusion and glucose hypometabolism: Key pathophysiological modulators promote neurodegeneration, cognitive impairment, and Alzheimer's disease. **Journal of neuroscience research**, Wiley Online Library, v. 95, n. 4, p. 943–972, 2017. DOI: <https://doi.org/10.1002/jnr.23777>. Citado na p. 15.

DAZE, K.; HOF, F. Molecular interaction and recognition. **Encyclopedia of physical organic chemistry**, Wiley Online Library, p. 1–51, 2016. DOI: <https://doi.org/10.1002/9781118468586.epoc3001>. Citado na p. 29.

DE PAULA, A.; MARTINS, J. B. L.; DOS SANTOS, M.; NASCENTE, L. d. C.; ROMEIRO, L. S.; VIEIRA, K. S. T.; GAMBÔA, N. F.; CASTRO, N. G. d.; GARGANO, R. New potential AChE inhibitor candidates. **European Journal of Medicinal Chemistry**, Elsevier, v. 44, n. 9, p. 3754–3759, 2009. Citado na p. 48.

DESIRAJU, G. R.; STEINER, T. **The weak hydrogen bond: in structural chemistry and biology**. [S. l.]: International Union of Crystal, 2001. v. 9. Citado na p. 29.

DESIRAJU, G. R.; VITTAL, J. J.; RAMANAN, A. **Crystal engineering: a textbook**. [S. l.]: World Scientific Publishing Company, 2011. Citado na p. 30.

DHARAMPAL, V. M. Methods of image edge detection: A review. **J. Electr. Electron. Syst**, v. 4, n. 2, p. 2332–0796, 2015. Citado na p. 40.

DOMINGO, L. R.; RÍOS-GUTIÉRREZ, M.; PÉREZ, P. Applications of the conceptual density functional theory indices to organic chemistry reactivity. **Molecules**, MDPI, v. 21, n. 6, p. 748, 2016. Citado na p. 29.

DOUMAR, H.; EL MOSTAFI, H.; ELHESSNI, A.; TOUHAMI, M. E.; MESFIOUI, A. Exploring the diversity of cannabis cannabinoid and non-cannabinoid compounds and their roles in Alzheimer's disease: A review. **IBRO Neuroscience Reports**, Elsevier, v. 18, p. 96–119, 2025. Citado na p. 21.

AL-EDRESI, S.; ALSALAHAT, I.; FREEMAN, S.; AOJULA, H.; PENNY, J. Resveratrol-mediated cleavage of amyloid β 1–42 peptide: potential relevance to Alzheimer's disease. **Neurobiology of Aging**, Elsevier, v. 94, p. 24–33, 2020. Citado na p. 26.

ELDUFANI, J.; BLAISE, G. The role of acetylcholinesterase inhibitors such as neostigmine and rivastigmine on chronic pain and cognitive function in aging: A review of recent clinical applications. **Alzheimer's & Dementia: Translational Research & Clinical Interventions**, Elsevier, v. 5, p. 175–183, 2019. Citado na p. 17.

ELVITARIA, L.; SHAUBARI, A.; FADZRIN, E.; SAMSUDIN, N. A.; KHALID, A.; KAMAL, S.; INDRA, Z. *et al.* A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method. **International Journal of Advanced Computer Science & Applications**, v. 15, n. 10, 2024. Citado na p. 39.

- ESTEVA, A.; CHOU, K.; YEUNG, S.; NAIK, N.; MADANI, A.; MOTTAGHI, A.; LIU, Y.; TOPOL, E.; DEAN, J.; SOCHER, R. Deep learning-enabled medical computer vision. **NPJ digital medicine**, Nature Publishing Group UK London, v. 4, n. 1, p. 5, 2021. Citado na p. 39.
- FÁ, M.; PUZZO, D.; PIACENTINI, R.; STANISZEWSKI, A.; ZHANG, H.; BALTRONS, M. A.; LI PUMA, D. D.; CHATTERJEE, I.; LI, J.; SAEED, F. *et al.* Extracellular tau oligomers produce an immediate impairment of LTP and memory. **Scientific reports**, Nature Publishing Group UK London, v. 6, n. 1, p. 19393, 2016. DOI: <https://doi.org/10.1038/srep19393>. Citado na p. 14.
- FANG, C.; LI, C.; CABRERIZO, M.; BARRETO, A.; ANDRIAN, J.; LOEWENSTEIN, D.; DUARA, R.; ADJOUADI, M. A Gaussian discriminant analysis-based generative learning algorithm for the early diagnosis of mild cognitive impairment in Alzheimer's disease. *In: IEEE. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). [S. l.: s. n.], 2017. p. 538–542.* Citado na p. 37.
- FANG, C.; LI, C.; CABRERIZO, M.; BARRETO, A.; ANDRIAN, J.; LOEWENSTEIN, D.; DUARA, R.; ADJOUADI, M. A novel Gaussian discriminant analysis-based computer aided diagnosis system for screening different stages of Alzheimer's disease. *In: IEEE. 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE). [S. l.: s. n.], 2017. p. 279–284.* Citado na p. 37.
- FANG, C.; LI, C.; FOROUZANNEZHAD, P.; CABRERIZO, M.; CURIEL, R. E.; LOEWENSTEIN, D.; DUARA, R.; ADJOUADI, M.; INITIATIVE, A. D. N. *et al.* Gaussian discriminative component analysis for early detection of Alzheimer's disease: A supervised dimensionality reduction algorithm. **Journal of neuroscience methods**, Elsevier, v. 344, p. 108856, 2020. Citado na p. 37.
- FANG, L.; YANG, T.; MEDINA-BOLIVAR, F. Production of prenylated stilbenoids in hairy root cultures of peanut (*Arachis hypogaea*) and its wild relatives *A. ipaensis* and *A. duranensis* via an optimized elicitation procedure. **Molecules**, MDPI, v. 25, n. 3, p. 509, 2020. Citado na p. 26.
- FERNÁNDEZ, J. A. Exploring hydrogen bond in biological molecules. **Journal of the Indian Institute of Science**, Springer, v. 100, n. 1, p. 135–154, 2020. DOI: <https://doi.org/10.1007/s41745-019-00146-4>. Citado na p. 29.
- FERNÁNDEZ-RUIZ, J. The biomedical challenge of neurodegenerative disorders: an opportunity for cannabinoid-based therapies to improve on the poor current therapeutic outcomes. **British journal of pharmacology**, Wiley Online Library, v. 176, n. 10, p. 1370–1383, 2019. DOI: <https://doi.org/10.1111/bph.14382>. Citado na p. 22.
- FINK, T.; BRUGGESSER, H.; REYMOND, J.-L. Virtual exploration of the small-molecule chemical universe below 160 daltons. **Angewandte Chemie International Edition**, Wiley Online Library, v. 44, n. 10, p. 1504–1508, 2005. Citado na p. 54.
- FINK, T.; REYMOND, J.-L. Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery. **Journal of chemical information and modeling**, ACS Publications, v. 47, n. 2, p. 342–353, 2007. Citado na p. 54.
- FISCHER, J.; GANELLIN, C. R. Analogue-based drug discovery. **Chemistry International–Newsmagazine for IUPAC**, De Gruyter, v. 32, n. 4, p. 12–15, 2010. Citado na p. 17.
- FORLI, S.; HUEY, R.; PIQUE, M. E.; SANNER, M. F.; GOODSSELL, D. S.; OLSON, A. J. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. **Nature protocols**, Nature Publishing Group UK London, v. 11, n. 5, p. 905–919, 2016. Citado na p. 31.
- FRASER, T. R. XLVIII.—On the Physiological Action of the Calabar Bean (*Physostigma venenosum*, Balf.) **Earth and Environmental Science Transactions of The Royal Society of Edinburgh**,

Royal Society of Edinburgh Scotland Foundation, v. 24, n. 3, p. 715–787, 1867. DOI: <https://doi.org/10.1017/s0080456800020585>. Citado na p. 18.

FRISCH, M. e.; TRUCKS, G.; SCHLEGEL, H. B.; SCUSERIA, G.; ROBB, M.; CHEESEMAN, J.; SCALMANI, G.; BARONE, V.; PETERSSON, G.; NAKATSUJI, H. *et al.* **Gaussian 16**. [S. l.]: Gaussian, Inc. Wallingford, CT, 2016. Citado na p. 49.

GALASKO, D.; COREY-BLOOM, J.; THAL, L. J. Monitoring progression in Alzheimer's disease. **Journal of the American Geriatrics Society**, Wiley Online Library, v. 39, n. 9, p. 932–941, 1991. Citado na p. 13.

GALVE-ROPERH, I.; PALAZUELOS, J.; AGUADO, T.; GUZMÁN, M. The endocannabinoid system and the regulation of neural development: potential implications in psychiatric disorders. **European archives of psychiatry and clinical neuroscience**, Springer, v. 259, p. 371–382, 2009. DOI: <https://doi.org/10.1007/s00406-009-0028-y>. Citado na p. 21.

GAO, Z.; WANG, L.; WU, G. Lip: Local importance-based pooling. *In*: PROCEEDINGS of the IEEE/CVF International Conference on Computer Vision. [S. l.: s. n.], 2019. p. 3355–3364. Citado na p. 46.

GEERLINGS, P.; DE PROFT, F.; LANGENAEKER, W. Conceptual density functional theory. **Chemical reviews**, ACS Publications, v. 103, n. 5, p. 1793–1874, 2003. Citado na p. 28.

GILBERT, G. L.; KIM, H. J.; WAATAJA, J. J.; THAYER, S. A. Δ^9 -Tetrahydrocannabinol protects hippocampal neurons from excitotoxicity. **Brain research**, Elsevier, v. 1128, p. 61–69, 2007. DOI: <https://doi.org/10.1016/j.brainres.2006.03.011>. Citado na p. 21.

GIRCHA, A.; BOEV, A. S.; AVCHACIOV, K.; FEDICHEV, P.; FEDOROV, A. K. Training a discrete variational autoencoder for generative chemistry and drug design on a quantum annealer. **arXiv preprint arXiv:2108.11644**, 2021. Citado na p. 46.

GONCA, E.; DARICI, F. The effect of cannabidiol on ischemia/reperfusion-induced ventricular arrhythmias: the role of adenosine A1 receptors. **Journal of cardiovascular pharmacology and therapeutics**, SAGE Publications Sage CA: Los Angeles, CA, v. 20, n. 1, p. 76–83, 2015. DOI: <https://doi.org/10.1177/1074248414532013>. Citado na p. 22.

GONZALEZ, R. C. **Digital image processing**. [S. l.]: Pearson education india, 2009. Citado na p. 39.

GOOD, P. F.; PERL, D. P.; BIERER, L. M.; SCHMEIDLER, J. Selective accumulation of aluminum and iron in the neurofibrillary tangles of Alzheimer's disease: a laser microprobe (LAMMA) study. **Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society**, Wiley Online Library, v. 31, n. 3, p. 286–292, 1992. DOI: <https://doi.org/10.1002/ana.410310310>. Citado na p. 14.

GOTTWALD, M. D.; ROZANSKI, R. I. Rivastigmine, a brain-region selective acetylcholinesterase inhibitor for treating Alzheimer's disease: review and current status. **Expert opinion on investigational drugs**, Taylor & Francis, v. 8, n. 10, p. 1673–1682, 1999. Citado na p. 17.

GOYAL, D.; KAUR, A.; GOYAL, B. Benzofuran and indole: promising scaffolds for drug development in Alzheimer's disease. **ChemMedChem**, Wiley Online Library, v. 13, n. 13, p. 1275–1299, 2018. DOI: <https://doi.org/10.1002/cmdc.201800156>. Citado na p. 25.

GUPTA, S. K.; KUMAR, V. S.; KHANG, A.; HAZELA, B.; NIVETHITHA, T.; HARALAYYA, B. Detection of Lung Tumor using an efficient Quadratic Discriminant Analysis Model. *In*: IEEE. 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC). [S. l.: s. n.], 2023. p. 1–6. Citado na p. 37.

HANANTO, A. L.; ASSIROJ, P.; PRIYATNA, B.; FAUZI, A.; RAHMAN, A. Y.; HILABI, S. S. *et al.* Analysis of drug data mining with clustering technique using K-Means algorithm. *In: IOP PUBLISHING, 1. JOURNAL of Physics: Conference Series*. [S. l.: s. n.], 2021. v. 1908, p. 012024. Citado na p. 38.

HARALICK, R.; SHANMUGAM, K. Dinstein, 1973. Texture features for image classification. **IEEE Trans. systems, man. & cibernetics**, **SMC-3**, p. 610–622. Citado na p. 39.

HARDY, J.; ALLSOP, D. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. **Trends in pharmacological sciences**, Elsevier Current Trends, v. 12, p. 383–388, 1991. Citado na p. 16.

HARINANTENAINA, L.; TAKAHARA, Y.; NISHIZAWA, T.; KOHCHI, C.; SOMA, G.-I.; ASAKAWA, Y. Chemical constituents of malagasy liverworts, Part V: Prenyl bibenzyls and clerodane diterpenoids with nitric oxide inhibitory activity from *Radula appressa* and *Thysananthus spathulistipus*. **Chemical and pharmaceutical bulletin**, The Pharmaceutical Society of Japan, v. 54, n. 7, p. 1046–1049, 2006. Citado na p. 23.

HARRIS, C. R.; MILLMAN, K. J.; VAN DER WALT, S. J.; GOMMERS, R.; VIRTANEN, P.; COUNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J. *et al.* Array programming with NumPy. **nature**, Nature Publishing Group UK London, v. 585, n. 7825, p. 357–362, 2020. Citado na p. 49.

HASAN, M. M.; CHA, M.; BAJPAI, V. K.; BAEK, K.-H. Production of a major stilbene phytoalexin, resveratrol in peanut (*Arachis hypogaea*) and peanut products: a mini review. **Reviews in Environmental Science and Bio/Technology**, Springer, v. 12, p. 209–221, 2013. Citado na p. 26.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *et al.* **The elements of statistical learning**. [S. l.]: Citeseer, 2009. Citado nas pp. 40, 41.

HAYKIN, S. **Redes neurais: princípios e prática**. [S. l.]: Bookman Editora, 2001. Citado na p. 35.

HEBB, D. O. Organization of behavior. new york: Wiley. **J. Clin. Psychol**, v. 6, n. 3, p. 335–307, 1949. Citado na p. 44.

HEMALATHA, B.; RENUKADEVI, M. Study of Alzheimer Disease Prediction Using Various Classifiers. **Annals of the Romanian Society for Cell Biology**, v. 25, n. 6, p. 10244–10250, 2021. Citado na p. 13.

HENRIQUEZ, J. E.; BACH, A. P.; MATOS-FERNANDEZ, K. M.; CRAWFORD, R. B.; KAMINSKI, N. E. Δ 9-Tetrahydrocannabinol (THC) Impairs CD8+ T Cell-Mediated Activation of Astrocytes. **Journal of Neuroimmune Pharmacology**, Springer, v. 15, p. 863–874, 2020. DOI: <https://doi.org/10.1007/s11481-020-09912-z>. Citado na p. 21.

HENSLEY, K.; CARNEY, J.; MATTSON, M.; AKSENOVA, M.; HARRIS, M.; WU, J.; FLOYD, R.; BUTTERFIELD, D. A model for beta-amyloid aggregation and neurotoxicity based on free radical generation by the peptide: relevance to Alzheimer disease. **Proceedings of the National Academy of Sciences**, v. 91, n. 8, p. 3270–3274, 1994. DOI: <https://doi.org/10.1073/pnas.91.8.3270>. Citado na p. 14.

HICKEY, J. P.; COLLINS, A. E.; NELSON, M. L.; CHEN, H.; KALISCH, B. E. Modulation of oxidative stress and Neuroinflammation by Cannabidiol (CBD): promising targets for the treatment of Alzheimer's disease. **Current Issues in Molecular Biology**, MDPI, v. 46, n. 5, p. 4379–4402, 2024. DOI: <https://doi.org/10.3390/cimb46050266>. Citado na p. 22.

HILGERT, S. S.; DIAS, D. P. M. The intestinal microbiota as an ally in the treatment of Alzheimer's disease. **Gut Microbiome**, Cambridge University Press, v. 4, e9, 2023. DOI: <https://doi.org/10.1017/gmb.2023.8>. Citado na p. 15.

HLAVCHEVA, D.; YALOVEHA, V.; PODOROZHNIAK, A. Application of convolutional neural network for histopathological analysis. **Advanced Information Systems**, v. 3, n. 4, p. 69–73, 2019. Citado na p. 47.

HOHENBERG, P.; KOHN, W. Density functional theory (DFT). **Phys. Rev.**, v. 136, n. 1964, b864, 1964. Citado na p. 28.

HOLTZMAN, A.; BUYS, J.; DU, L.; FORBES, M.; CHOI, Y. The curious case of neural text degeneration. **arXiv preprint arXiv:1904.09751**, 2019. Citado na p. 72.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 79, n. 8, p. 2554–2558, 1982. Citado na p. 44.

HU, C.; LI, S.; YANG, C.; CHEN, J.; XIONG, Y.; FAN, G.; LIU, H.; HONG, L. ScaffoldGVAE: scaffold generation and hopping of drug molecules via a variational autoencoder based on multi-view graph neural networks. **Journal of Cheminformatics**, Springer, v. 15, n. 1, p. 91, 2023. Citado na p. 46.

HU, T.; CHEN, F.; WANG, H.; LI, J.; WANG, W.; SUN, J.; LI, Z. Complexity matters: Rethinking the latent space for generative modeling. **Advances in Neural Information Processing Systems**, v. 36, p. 29558–29579, 2023. Citado na p. 72.

HUANG, Q.; LIAO, C.; GE, F.; AO, J.; LIU, T. Acetylcholine bidirectionally regulates learning and memory. **Journal of Neurorestoratology**, Elsevier, v. 10, n. 2, p. 100002, 2022. Citado na p. 15.

HUANG, Z.; WONG, L.-W.; SU, Y.; HUANG, X.; WANG, N.; CHEN, H.; YI, C. Blood-brain barrier integrity in the pathogenesis of Alzheimer’s disease. **Frontiers in neuroendocrinology**, Elsevier, v. 59, p. 100857, 2020. DOI: <https://doi.org/10.1016/j.yfrne.2020.100857>. Citado na p. 15.

HUANG, Z.; CHEN, S.; YU, L. Predicting new drug indications based on double variational autoencoders. **Computers in Biology and Medicine**, Elsevier, v. 164, p. 107261, 2023. Citado na p. 46.

HUDSON, S.; RAMSEY, J.; KING, L.; TIMBERS, S.; MAYNARD, S.; DARGAN, P. I.; WOOD, D. M. Use of high-resolution accurate mass spectrometry to detect reported and previously unreported cannabinomimetics in “herbal high” products. **Journal of analytical toxicology**, Oxford University Press, v. 34, n. 5, p. 252–260, 2010. DOI: <https://doi.org/10.1093/jat/34.5.252>. Citado na p. 22.

HUGHES, J. A.; PERSKY, A. M.; GUO, X.; GHOSH, T. K. Solubility. *In: THEORY and Practice of Contemporary Pharmaceutics*. [S. l.]: CRC press, 2021. p. 55–81. DOI: <https://doi.org/10.1201/9780203644478-4>. Citado na p. 29.

HUNTER, C. A.; SANDERS, J. K. The nature of pi-...pi interactions. **Journal of the American Chemical Society**, ACS Publications, v. 112, n. 14, p. 5525–5534, 1990. Citado na p. 30.

HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007. DOI: 10.1109/MCSE.2007.55. Citado na p. 49.

IMRIE, B.; JONES, R.; KERRIDGE, P. Desmodium. *In: THE role of Centrosema, Desmodium, and Stylosanthes in improving tropical pastures*. [S. l.]: CRC Press, 2019. p. 97–140. Citado na p. 25.

IONESCU-TUCKER, A.; COTMAN, C. W. Emerging roles of oxidative stress in brain aging and Alzheimer’s disease. **Neurobiology of aging**, Elsevier, v. 107, p. 86–95, 2021. DOI: <https://doi.org/10.1016/j.neurobiolaging.2021.07.014>. Citado na p. 14.

IQBAL, K.; GONG, C.-X.; LIU, F. Hyperphosphorylation-induced tau oligomers. **Frontiers in neurology**, Frontiers Media SA, v. 4, p. 112, 2013. DOI: <https://doi.org/10.3389/fneur.2013.00112>. Citado na p. 14.

- IRWIN, J. J.; SHOICHET, B. K. ZINC- a free database of commercially available compounds for virtual screening. **Journal of chemical information and modeling**, ACS Publications, v. 45, n. 1, p. 177–182, 2005. Citado na p. 54.
- ISLAM, M. R.; ZAMIL, M. Z. H.; RAYED, M. E.; KABIR, M. M.; MRIDHA, M.; NISHIMURA, S.; SHIN, J. Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. **IEEE Access**, IEEE, 2024. Citado na p. 39.
- ISRAELACHVILI, J. N. **Intermolecular and surface forces**. [S. l.]: Academic press, 2011. Citado na p. 30.
- JACK, M.; ERIKS, K. **Adamantyl secondary amines**. [S. l.]: Google Patents, jul. 1968. US Patent 3,391,142. Citado na p. 20.
- JANAI, J.; GÜNEY, F.; BEHL, A.; GEIGER, A. *et al.* Computer vision for autonomous vehicles: Problems, datasets and state of the art. **Foundations and Trends® in Computer Graphics and Vision**, Now Publishers, Inc., v. 12, n. 1–3, p. 1–308, 2020. Citado na p. 39.
- JANN, M. W. Rivastigmine, a new-generation cholinesterase inhibitor for the treatment of Alzheimer's disease. **Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy**, Wiley Online Library, v. 20, n. 1, p. 1–12, 2000. Citado na p. 17.
- JARROTT, B. Tacrine: In vivo veritas. **Pharmacological research**, Elsevier, v. 116, p. 29–31, 2017. DOI: <https://doi.org/10.1016/j.phrs.2016.12.033>. Citado na p. 17.
- JHA, S. K.; NELSON, V. K.; SURYADEVARA, P. R.; PANDA, S. P.; PULLAIAH, C. P.; NULI, M. V.; KAMAL, M.; IMRAN, M.; AUSALI, S.; ABOMUGHAID, M. M. *et al.* Cannabidiol and neurodegeneration: From molecular mechanisms to clinical benefits. **Ageing research reviews**, Elsevier, v. 100, p. 102386, 2024. DOI: <https://doi.org/10.1016/j.arr.2024.102386>. Citado na p. 22.
- JOHN, H. **Adaptation in natural and artificial systems**. [S. l.: s. n.], 2019. Citado na p. 67.
- JOHNSON, A. P.; BAKER JR, C. L. First-and second-order information in natural images: a filter-based approach to image statistics. **Journal of the Optical Society of America A**, Optical Society of America, v. 21, n. 6, p. 913–925, 2004. Citado na p. 39.
- JOHNSON, E. R.; KEINAN, S.; MORI-SÁNCHEZ, P.; CONTRERAS-GARCÍA, J.; COHEN, A. J.; YANG, W. Revealing noncovalent interactions. **Journal of the American Chemical Society**, ACS Publications, v. 132, n. 18, p. 6498–6506, 2010. Citado na p. 30.
- JOLLIFFE, I. T. **Principal component analysis for special types of data**. [S. l.]: Springer, 2002. Citado nas pp. 40, 43.
- JOLY, L.; GUERRA, S.; SÉPTIMO, R.; SOLÍS, P.; CORREA, M.; GUPTA, M.; LEVY, S.; SANDBERG, F. Ethnobotanical inventory of medicinal plants used by the Guaymí Indians in Western Panama. Part I. **Journal of Ethnopharmacology**, Elsevier, v. 20, n. 2, p. 145–171, 1987. DOI: [https://doi.org/10.1016/0378-8741\(87\)90085-7](https://doi.org/10.1016/0378-8741(87)90085-7). Citado na p. 24.
- KARNILA, S.; IRIANTO, S. Face recognition using content based image retrieval for intelligent security. **Intelligent Retrieval**, v. 2, n. 2, p. 66–69, 2024. Citado na p. 39.
- KASATKINA, L. A.; RITTCHEN, S.; STURM, E. M. Neuroprotective and immunomodulatory action of the endocannabinoid system under neuroinflammation. **International Journal of Molecular Sciences**, MDPI, v. 22, n. 11, p. 5431, 2021. DOI: <https://doi.org/10.3390/ijms22115431>. Citado na p. 21.

KEILUWEIT, M.; KLEBER, M. Molecular-level interactions in soils and sediments: the role of aromatic π -systems. **Environmental science & technology**, ACS Publications, v. 43, n. 10, p. 3421–3429, 2009. Citado na p. 30.

KLETTE, R. **Concise computer vision**. [S. l.]: Springer, 2014. v. 233. Citado na p. 39.

KUMAR, R.; SHARMA, A.; VARADWAJ, P. K. A prediction model for oral bioavailability of drugs using physicochemical properties by support vector machine. **Journal of natural science, biology, and medicine**, Wolters Kluwer–Medknow Publications, v. 2, n. 2, p. 168, 2011. Citado na p. 32.

KUMARIHAMY, M.; TRIPATHI, S.; BALACHANDRAN, P.; AVULA, B.; ZHAO, J.; WANG, M.; BENNETT, M. M.; ZHANG, J.; CARR, M. A.; LOVELL, K. M. *et al.* Synthesis and inhibitory activity of machaeridiol-based novel anti-MRSA and anti-VRE compounds and their profiling for cancer-related signaling pathways. **Molecules**, MDPI, v. 27, n. 19, p. 6604, 2022. DOI: <https://doi.org/10.3390/molecules27196604>. Citado na p. 24.

KUME, T.; TAKADA-TAKATORI, Y. Nicotinic acetylcholine receptor signaling: Roles in neuroprotection. **Nicotinic acetylcholine receptor signaling in neuroprotection**, Springer Singapore, p. 59–71, 2018. Citado na p. 17.

KUNG, S. Y. **Kernel methods and machine learning**. [S. l.]: Cambridge University Press, 2014. Citado na p. 37.

LATTIBEAUDIERE, K.; NEWMAN, S. A Review of the Multifaceted Plant, *Desmodium canum*: The Traditional and Ethnopharmacological Applications and its Potential Use in Medicine. **Pharmacognosy Reviews**, v. 18, n. 35, 2024. Citado na p. 25.

LEE, A. Lipid–protein interactions in biological membranes: a structural perspective. **Biochimica et Biophysica Acta (BBA)-Biomembranes**, Elsevier, v. 1612, n. 1, p. 1–40, 2003. DOI: [https://doi.org/10.1016/s0005-2736\(03\)00056-7](https://doi.org/10.1016/s0005-2736(03)00056-7). Citado na p. 30.

LEE, C.; YANG, W.; PARR, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Physical review B**, APS, v. 37, n. 2, p. 785, 1988. Citado na p. 49.

LEMES, L. F. N.; ANDRADE RAMOS, G. de; DE OLIVEIRA, A. S.; DA SILVA, F. M. R.; CASTRO COUTO, G. de; SILVA BONI, M. da; GUIMARÃES, M. J. R.; SOUZA, I. N. O.; BARTOLINI, M.; ANDRISANO, V. *et al.* Cardanol-derived AChE inhibitors: Towards the development of dual binding derivatives for Alzheimer’s disease. **European Journal of Medicinal Chemistry**, Elsevier, v. 108, p. 687–700, 2016. DOI: <https://doi.org/10.1016/j.ejmech.2015.12.024>. Citado na p. 18.

LI, H.; YU, S.; PRINCIPE, J. Causal recurrent variational autoencoder for medical time series generation. *In*: 7. PROCEEDINGS of the AAAI conference on artificial intelligence. [S. l.: s. n.], 2023. v. 37, p. 8562–8570. Citado na p. 46.

LI, Q.; CHEN, L.; LIU, X.; LI, X.; CAO, Y.; BAI, Y.; QI, F. Pterostilbene inhibits amyloid- β -induced neuroinflammation in a microglia cell line by inactivating the NLRP3/caspase-1 inflammasome pathway. **Journal of cellular biochemistry**, Wiley Online Library, v. 119, n. 8, p. 7053–7062, 2018. Citado na p. 26.

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE transactions on neural networks and learning systems**, IEEE, v. 33, n. 12, p. 6999–7019, 2021. Citado na p. 46.

LIM, T. K. *Anacardium occidentale*. *In*: EDIBLE Medicinal and Non-Medicinal Plants: Volume 1, Fruits. [S. l.]: Springer, 2011. p. 45–68. DOI: https://doi.org/10.1007/978-90-481-8661-7_6. Citado na p. 18.

- LIN, J.; HUANG, L.; YU, J.; XIANG, S.; WANG, J.; ZHANG, J.; YAN, X.; CUI, W.; HE, S.; WANG, Q. Fucoxanthin, a marine carotenoid, reverses scopolamine-induced cognitive impairments in mice and inhibits acetylcholinesterase in vitro. **Marine Drugs**, MDPI, v. 14, n. 4, p. 67, 2016. DOI: <https://doi.org/10.3390/md14040067>. Citado na p. 20.
- LINNAINMAA, S. **The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors**. 1970. Tese (Doutorado) – Master's Thesis (in Finnish), Univ. Helsinki. Citado na p. 45.
- LIU, L.; SHEN, C.; VAN DEN HENGEL, A. The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. *In*: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition. [S. l.: s. n.], 2015. p. 4749–4757. Citado na p. 46.
- LIU, P.-P.; XIE, Y.; MENG, X.-Y.; KANG, J.-S. History and progress of hypotheses and clinical trials for Alzheimer's disease. **Signal transduction and targeted therapy**, Nature Publishing Group UK London, v. 4, n. 1, p. 29, 2019. Citado na p. 16.
- LLOYD, S. Least square quantization in PCM. Bell Telephone Laboratories Paper. Published in journal much later: Lloyd, SP: Least squares quantization in PCM. **IEEE Trans. Inform. Theor.**(1957/1982), v. 18, n. 11, 1957. Citado na p. 38.
- LOEWI, O. Über humorale übertragbarkeit der Herznervenwirkung. **Pflügers Archiv European Journal of Physiology**, Springer Berlin/Heidelberg, v. 189, n. 1, p. 239–242, 1921. Citado na p. 15.
- LORIMER, S. D.; PERRY, N. B.; TANGNEY, R. S. An antifungal bibenzyl from the New Zealand liverwort, *Plagiochila stephensoniana*. Bioactivity-directed isolation, synthesis, and analysis. **Journal of natural products**, ACS Publications, v. 56, n. 9, p. 1444–1450, 1993. Citado na p. 22.
- LOWE, H.; TOYANG, N.; STEELE, B.; BRYANT, J.; NGWA, W. The endocannabinoid system: a potential target for the treatment of various diseases. **International journal of molecular sciences**, MDPI, v. 22, n. 17, p. 9472, 2021. DOI: <https://doi.org/10.3390/ijms22179472>. Citado na p. 21.
- LU, J.; CHEN, L.; YIN, J.; HUANG, T.; BI, Y.; KONG, X.; ZHENG, M.; CAI, Y.-D. Identification of new candidate drugs for lung cancer using chemical–chemical interactions, chemical–protein interactions and a K-means clustering algorithm. **Journal of Biomolecular Structure and Dynamics**, Taylor & Francis, v. 34, n. 4, p. 906–917, 2016. Citado na p. 38.
- LU, T.; MULTIWFN, R. 2.6. 1; University of Science and Technology Beijing: Beijing, China, 2013;(b) T. Lu, F. Chen. **J. Comput. Chem**, v. 33, p. 580, 2012. Citado na p. 49.
- LUBIS, I.; SIREGAR, R. *et al.* Recognition of Medicinal Plant Leaf Patterns Using Morphology-Based and GLCM Feature Extraction. **Journal of Artificial Intelligence and Engineering Applications (JAIEA)**, v. 3, n. 3, p. 834–839, 2024. Citado na p. 39.
- MA, J. C.; DOUGHERTY, D. A. The cation- π interaction. **Chemical reviews**, ACS Publications, v. 97, n. 5, p. 1303–1324, 1997. Citado na p. 30.
- MABBERLEY, D. J. **Mabberley's plant-book: a portable dictionary of plants, their classification and uses**. [S. l.: s. n.], 2017. DOI: <https://doi.org/10.1017/9781316335581>. Citado na p. 24.
- MACQUEEN, J. *et al.* Some methods for classification and analysis of multivariate observations. *In*: OAKLAND, CA, USA, 14. PROCEEDINGS of the fifth Berkeley symposium on mathematical statistics and probability. [S. l.: s. n.], 1967. v. 1, p. 281–297. Citado na p. 38.
- MADUGULA, S. S.; JOHN, L.; NAGAMANI, S.; GAUR, A. S.; POROIKOV, V. V.; SASTRY, G. N. Molecular descriptor analysis of approved drugs using unsupervised learning for drug repurposing. **Computers in Biology and Medicine**, Elsevier, v. 138, p. 104856, 2021. Citado na p. 38.

MAHMOUD MOUSTAFA, E.; RASHED, E. R.; RASHED, R. R.; OMAR, N. N. Piceatannol promotes hepatic and renal AMPK/SIRT1/PGC-1 α mitochondrial pathway in rats exposed to reserpine or gamma-radiation. **International Journal of Immunopathology and Pharmacology**, SAGE Publications Sage UK: London, England, v. 35, p. 20587384211016194, 2021. Citado na p. 26.

MAK, S.; LI, W.; FU, H.; LUO, J.; CUI, W.; HU, S.; PANG, Y.; CARLIER, P. R.; TSIM, K. W.; PI, R. *et al.* Promising tacrine/huperzine A-based dimeric acetylcholinesterase inhibitors for neurodegenerative disorders: From relieving symptoms to modifying diseases through multitarget. **Journal of neurochemistry**, Wiley Online Library, v. 158, n. 6, p. 1381–1393, 2021. Citado na p. 18.

MALLET, V.; CHECA RUANO, L.; MOINE FRANEL, A.; NILGES, M.; DRUART, K.; BOUVIER, G.; SPERANDIO, O. InDeep: 3D fully convolutional neural networks to assist in silico drug design on protein–protein interactions. **Bioinformatics**, Oxford University Press, v. 38, n. 5, p. 1261–1268, 2022. Citado na p. 47.

MARENICH, A. V.; CRAMER, C. J.; TRUHLAR, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. **The Journal of Physical Chemistry B**, ACS Publications, v. 113, n. 18, p. 6378–6396, 2009. Citado na p. 49.

MAROON, J.; BOST, J. Review of the neurological benefits of phytocannabinoids. **Surgical neurology international**, v. 9, p. 91, 2018. DOI: https://doi.org/10.4103/sni.sni_45_18. Citado na p. 21.

MARTÍN ABADI; ASHISH AGARWAL; PAUL BARHAM; EUGENE BREVDO; ZHIFENG CHEN; CRAIG CITRO; GREG S. CORRADO; ANDY DAVIS; JEFFREY DEAN; MATTHIEU DEVIN; SANJAY GHEMAWAT; IAN GOODFELLOW; ANDREW HARP; GEOFFREY IRVING; MICHAEL ISARD; JIA, Y.; RAFAL JOZEFOWICZ; LUKASZ KAISER; MANJUNATH KUDLUR; JOSH LEVENBERG; DANDELION MANÉ; RAJAT MONGA; SHERRY MOORE; DEREK MURRAY; CHRIS OLAH; MIKE SCHUSTER; JONATHON SHLENS; BENOIT STEINER; ILYA SUTSKEVER; KUNAL TALWAR; PAUL TUCKER; VINCENT VANHOUCHE; VIJAY VASUDEVAN; FERNANDA VIÉGAS; ORIOL VINYALS; PETE WARDEN; MARTIN WATTENBERG; MARTIN WICKE; YUAN YU; XIAOQIANG ZHENG. **TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems**. [S. l.: s. n.], 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>. Citado na p. 50.

MARTYN, C.; OSMOND, C.; EDWARDSON, J.; BARKER, D.; HARRIS, E.; LACEY, R. Geographical relation between Alzheimer's disease and aluminium in drinking water. **The Lancet**, Elsevier, v. 333, n. 8629, p. 59–62, 1989. DOI: [https://doi.org/10.1016/s0140-6736\(89\)91425-6](https://doi.org/10.1016/s0140-6736(89)91425-6). Citado na p. 14.

MARUCCI, G.; BUCCIONI, M.; DAL BEN, D.; LAMBERTUCCI, C.; VOLPINI, R.; AMENTA, F. Efficacy of acetylcholinesterase inhibitors in Alzheimer's disease. **Neuropharmacology**, Elsevier, v. 190, p. 108352, 2021. Citado na p. 17.

MASTERS, C. L.; BATEMAN, R.; BLENNOW, K.; ROWE, C. C.; SPERLING, R. A.; CUMMINGS, J. L. Alzheimer's disease. **Nature reviews disease primers**, Nature Publishing Group, v. 1, n. 1, p. 1–18, 2015. Citado nas pp. 13, 16.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, p. 115–133, 1943. Citado na p. 44.

MCKINNEY, W. Data Structures for Statistical Computing in Python. *In*: WALT, S. van der; MILLMAN, J. (ed.). **Proceedings of the 9th Python in Science Conference**. [S. l.: s. n.], 2010. p. 56–61. DOI: 10.25080/Majora-92bf1922-00a. Citado na p. 49.

- MEDA, L.; BARON, P.; SCARLATO, G. Glial activation in Alzheimer's disease: the role of A β and its associated proteins. **Neurobiology of aging**, Elsevier, v. 22, n. 6, p. 885–893, 2001. DOI: [https://doi.org/10.1016/s0197-4580\(01\)00307-4](https://doi.org/10.1016/s0197-4580(01)00307-4). Citado na p. 14.
- MEDEIROS, R.; LAFERLA, F. M. Astrocytes: conductors of the Alzheimer disease neuroinflammatory symphony. **Experimental neurology**, Elsevier, v. 239, p. 133–138, 2013. DOI: <https://doi.org/10.1016/j.expneurol.2012.10.007>. Citado na p. 14.
- MEHMOOD, A.; MAQSOOD, M.; BASHIR, M.; SHUYUAN, Y. A deep Siamese convolution neural network for multi-class classification of Alzheimer disease. **Brain sciences**, MDPI, v. 10, n. 2, p. 84, 2020. Citado na p. 13.
- MEI, H.; ZHOU, Y.; LIANG, G.; LI, Z. Support vector machine applied in QSAR modelling. **Chinese Science Bulletin**, Springer, v. 50, p. 2291–2296, 2005. Citado na p. 32.
- MERCER, J. Xvi. functions of positive and negative type, and their connection the theory of integral equations. **Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character**, The Royal Society London, v. 209, n. 441–458, p. 415–446, 1909. Citado na p. 36.
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The Journal of chemical physics**, American Institute of Physics, v. 21, n. 6, p. 1087–1092, 1953. Citado na p. 67.
- MEVERS, E.; MATAINAH, T.; ALLARA', M.; DI MARZO, V.; GERWICK, W. H. Mooreamide A: A cannabinomimetic lipid from the marine cyanobacterium *Moorea bouillonii*. **Lipids**, Springer, v. 49, p. 1127–1132, 2014. DOI: <https://doi.org/10.1007/s11745-014-3949-9>. Citado na p. 22.
- MIRAKHORI, F.; MOAFI, M.; MILANIFARD, M.; TAHERNIA, H. *et al.* Diagnosis and Treatment Methods in Alzheimer's Patients Based on Modern Techniques: The Original Article. **Journal of Pharmaceutical Negative Results**, p. 1889–1907, 2022. Citado na p. 13.
- MISHRA, C. B.; KUMARI, S.; MANRAL, A.; PRAKASH, A.; SAINI, V.; LYNN, A. M.; TIWARI, M. Design, synthesis, in-silico and biological evaluation of novel donepezil derivatives as multi-target-directed ligands for the treatment of Alzheimer's disease. **European journal of medicinal chemistry**, Elsevier, v. 125, p. 736–750, 2017. Citado na p. 19.
- MITCHELL, T. M.; MITCHELL, T. M. **Machine learning**. [S. l.]: McGraw-hill New York, 1997. v. 1. Citado na p. 35.
- MIYAZAKI, T.; TAKENAKA, T.; INOUE, T.; SATO, M.; MIYAJIMA, Y.; NODERA, M.; HANYU, M.; OHNO, Y.; SHIBAZAKI, S.; SUZUKI, H. Lipopolysaccharide-induced overproduction of nitric oxide and overexpression of iNOS and interleukin-1 β proteins in zinc-deficient rats. **Biological trace element research**, Springer, v. 145, p. 375–381, 2012. Citado na p. 23.
- MIYAZAWA, M.; ASAKAWA, Y. Symposium on the Chemistry of Terpenes, Essential Oils and Aromatics (TEAC). **Journal of Oleo Science**, Japan Oil Chemists' Society, v. 67, n. 10, p. 1177–1177, 2018. Citado na p. 23.
- MONTERO, J. M. S.; AGIS-TORRES, A.; SOLANO, D.; SÖLLHUBER, M.; FERNANDEZ, M.; VILLARO, W.; GÓMEZ-CAÑAS, M.; GARCÍA-ARENCIBIA, M.; FERNÁNDEZ-RUIZ, J.; EGEA, J. *et al.* Analogues of cannabinoids as multitarget drugs in the treatment of Alzheimer's disease. **European Journal of Pharmacology**, Elsevier, v. 895, p. 173875, 2021. DOI: <https://doi.org/10.1016/j.ejphar.2021.173875>. Citado na p. 21.
- MORRIS, G. M.; GOODSELL, D. S.; HALLIDAY, R. S.; HUEY, R.; HART, W. E.; BELEW, R. K.; OLSON, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free

energy function. **Journal of computational chemistry**, Wiley Online Library, v. 19, n. 14, p. 1639–1662, 1998. Citado na p. 48.

MORRIS, G. M.; GOODSELL, D. S.; HUEY, R.; HART, W. E.; HALLIDAY, S.; BELEW, R.; OLSON, A. J. AutoDock. **Automated docking of flexible ligands to receptor-User Guide**, Citeseer, 2001. Citado na p. 30.

MORRIS, G. M.; HUEY, R.; LINDSTROM, W.; SANNER, M. F.; BELEW, R. K.; GOODSELL, D. S.; OLSON, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. **Journal of computational chemistry**, Wiley Online Library, v. 30, n. 16, p. 2785–2791, 2009. Citado na p. 31.

MUHAMMAD, I.; IBRAHIM, M. A.; KUMARIHAMY, M.; LAMBERT, J. A.; ZHANG, J.; MOHAMMAD, M. H.; KHAN, S. I.; PASCO, D. S.; BALACHANDRAN, P. Cannabinoid and Opioid Receptor Affinity and Modulation of Cancer-Related Signaling Pathways of Machaeriols and Machaeridiols from Machaerium Pers. **Molecules**, MDPI, v. 28, n. 10, p. 4162, 2023. DOI: <https://doi.org/10.3390/molecules28104162>. Citado na p. 25.

MUHAMMAD, I.; JACOB, M. R.; IBRAHIM, M. A.; RAMAN, V.; KUMARIHAMY, M.; WANG, M.; AL-ADHAMI, T.; HIND, C.; CLIFFORD, M.; MARTIN, B. *et al.* Antimicrobial constituents from Machaerium Pers.: Inhibitory activities and synergism of machaeriols and machaeridiols against methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococcus faecium*, and permeabilized gram-negative pathogens. **Molecules**, MDPI, v. 25, n. 24, p. 6000, 2020. DOI: <https://doi.org/10.3390/molecules25246000>. Citado na p. 24.

MUHAMMAD, I.; LI, X.-C.; DUNBAR, D. C.; ELSOHLI, M. A.; KHAN, I. A. Antimalarial (+)-trans-hexahydrodibenzopyran derivatives from Machaerium multiflorum. **Journal of natural products**, ACS Publications, v. 64, n. 10, p. 1322–1325, 2001. DOI: <https://doi.org/10.1021/np0102861>. Citado na p. 24.

MUHAMMAD, I.; LI, X.-C.; JACOB, M. R.; TEKWANI, B. L.; DUNBAR, D. C.; FERREIRA, D. Antimicrobial and Antiparasitic (+)-trans-Hexahydrodibenzopyrans and Analogues from Machaerium multiflorum. **Journal of natural products**, ACS Publications, v. 66, n. 6, p. 804–809, 2003. DOI: <https://doi.org/10.1021/np030045o>. Citado na p. 24.

MUKESH, K.; IPPATAPU VENKATA, S.; CHEREDDY, S.; ANBAZHAGAN, E.; OVIYA, I. A variational autoencoder—general adversarial networks (vae-gan) based model for ligand designing. *In*: SPRINGER. INTERNATIONAL Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 1. [S. l.: s. n.], 2022. p. 761–768. Citado na p. 46.

MUMA, N.; SINGER, S. Aluminum-induced neuropathology: transient changes in microtubule-associated proteins. **Neurotoxicology and teratology**, Elsevier, v. 18, n. 6, p. 679–690, 1996. DOI: [https://doi.org/10.1016/s0892-0362\(96\)00126-2](https://doi.org/10.1016/s0892-0362(96)00126-2). Citado na p. 14.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. *In*: PROCEEDINGS of the 27th international conference on machine learning (ICML-10). [S. l.: s. n.], 2010. p. 807–814. Citado na p. 44.

NELSON, D. L.; COX, M. M. **Princípios de bioquímica de Lehninger**. [S. l.]: Artmed Editora, 2022. Citado na p. 30.

NERI, L.; HEWITT, D. Aluminium, Alzheimer's disease, and drinking water. **The Lancet**, Elsevier, v. 338, n. 8763, p. 390, 1991. DOI: [https://doi.org/10.1016/0140-6736\(91\)90531-s](https://doi.org/10.1016/0140-6736(91)90531-s). Citado na p. 14.

NG, A. Generative learning algorithms. **Stanford Lecture Notes**, v. 5, n. 4, 2008. Citado na p. 37.

- NOBEL. **The Nobel Prize in Physics 2024** — [nobelprize.org](https://www.nobelprize.org). [S. l.: s. n.], 2024. <https://www.nobelprize.org/prizes/physics/2024/summary/>. [Accessed 25-11-2024]. Citado na p. 44.
- OH, J. H.; CHOI, J. S.; NAM, T.-J. Fucosterol from an edible brown alga *Ecklonia stolonifera* prevents soluble amyloid beta-induced cognitive dysfunction in aging rats. **Marine Drugs**, MDPI, v. 16, n. 10, p. 368, 2018. DOI: <https://doi.org/10.3390/md16100368>. Citado na p. 20.
- OLSSON, M. H. M.; SØNDERGAARD, C. R.; ROSTKOWSKI, M.; JENSEN, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKaPredictions. **Journal of Chemical Theory and Computation**, American Chemical Society (ACS), v. 7, n. 2, p. 525–537, jan. 2011. ISSN 1549-9626. DOI: 10.1021/ct100578z. Disponível em: <http://dx.doi.org/10.1021/ct100578z>. Citado na p. 48.
- PARR, R. G.; DONNELLY, R. A.; LEVY, M.; PALKE, W. E. Electronegativity: the density functional viewpoint. **The Journal of chemical physics**, American Institute of Physics, v. 68, n. 8, p. 3801–3807, 1978. Citado na p. 28.
- PARSONS, C. G.; DANYSZ, W.; BARTMANN, A.; SPIELMANN, P.; FRANKIEWICZ, T.; HESSELINK, M.; EILBACHER, B.; QUACK, G. Amino-alkyl-cyclohexanes are novel uncompetitive NMDA receptor antagonists with strong voltage-dependency and fast blocking kinetics: in vitro and in vivo characterization. **Neuropharmacology**, Elsevier, v. 38, n. 1, p. 85–108, 1999. Citado na p. 20.
- PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L. *et al.* Pytorch: An imperative style, high-performance deep learning library. **Advances in neural information processing systems**, v. 32, 2019. Citado na p. 50.
- PATTANAYAK, S. K.; CHOWDHURI, S. Effect of water on solvation structure and dynamics of ions in the peptide bond environment: importance of hydrogen bonding and dynamics of the solvents. **The Journal of Physical Chemistry B**, ACS Publications, v. 115, n. 45, p. 13241–13252, 2011. DOI: <https://doi.org/10.1021/jp206027e>. Citado na p. 29.
- PEARSON, K. LIII. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin philosophical magazine and journal of science**, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901. Citado na p. 40.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. Citado na p. 49.
- PENG, J.; FAN, M.; AN, C.; NI, F.; HUANG, W.; LUO, J. A narrative review of molecular mechanism and therapeutic effect of cannabidiol (CBD). **Basic & clinical pharmacology & toxicology**, Wiley Online Library, v. 130, n. 4, p. 439–456, 2022. DOI: <https://doi.org/10.1111/bcpt.13710>. Citado na p. 22.
- PERDEW, J. P.; BURKE, K.; ERNZERHOF, M. Generalized gradient approximation made simple. **Physical review letters**, APS, v. 77, n. 18, p. 3865, 1996. Citado na p. 30.
- PIAZZI, L.; RAMPA, A.; BISI, A.; GOBBI, S.; BELLUTI, F.; CAVALLI, A.; BARTOLINI, M.; ANDRISANO, V.; VALENTI, P.; RECANATINI, M. 3-(4-([Benzyl (methyl) amino] methyl} phenyl)-6,7-dimethoxy-2 H-2-chromenone (AP2238) Inhibits Both Acetylcholinesterase and Acetylcholinesterase-Induced β -Amyloid Aggregation: A Dual Function Lead for Alzheimer's Disease Therapy. **Journal of medicinal chemistry**, ACS Publications, v. 46, n. 12, p. 2279–2282, 2003. DOI: <https://doi.org/10.1021/jm0340602>. Citado na p. 19.

POLHILL, R.; RAVEN, P. Advances in legume systematics, Royal Botanic Gardens, Kew and Ministry of Agriculture. **Fisheries and Food, London**, 1981. DOI: <https://doi.org/10.1126/science.216.4553.1402.b>. Citado na p. 24.

POLINSKY, R. J. Clinical pharmacology of rivastigmine: a new-generation acetylcholinesterase inhibitor for the treatment of Alzheimer's disease. **Clinical therapeutics**, Elsevier, v. 20, n. 4, p. 634–647, 1998. Citado na p. 17.

PRATT, R.; PERDOMO, C.; SURICK, I.; IENI, J. Donepezil: tolerability and safety in Alzheimer's disease. **International journal of clinical practice**, Wiley Online Library, v. 56, n. 9, p. 710–717, 2002. Citado na p. 17.

R DE MELLO SCHIER, A.; P DE OLIVEIRA RIBEIRO, N.; S COUTINHO, D.; MACHADO, S.; ARIAS-CARRIÓN, O.; A CRIPPA, J.; W ZUARDI, A.; E NARDI, A.; C SILVA, A. Antidepressant-like and anxiolytic-like effects of cannabidiol: a chemical compound of Cannabis sativa. **CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)**, Bentham Science Publishers, v. 13, n. 6, p. 953–960, 2014. DOI: <https://doi.org/10.2174/1871527313666140612114838>. Citado na p. 22.

RAFALOVSKA, E.; XHEMAI, E.; NAUMOVSKA, Z.; GIGOPULU, O.; MIOVA, B.; SUTURKOVA, L.; STEFKOV, G. Unlocking the Antidiabetic Potential of CBD: In Vivo Preclinical Studies. **Pharmaceuticals**, MDPI, v. 18, n. 4, p. 446, 2025. DOI: <https://doi.org/10.3390/ph18040446>. Citado na p. 22.

RAY, B.; MALONEY, B.; SAMBAMURTI, K.; KARNATI, H. K.; NELSON, P. T.; GREIG, N. H.; LAHIRI, D. K. Rivastigmine modifies the α -secretase pathway and potentially early Alzheimer's disease. **Translational psychiatry**, Nature Publishing Group UK London, v. 10, n. 1, p. 47, 2020. Citado na p. 17.

REHMAN, I.; KERNDT, C. C.; BOTELHO, S. Biochemistry, tertiary protein structure. *In*: STATPEARLS [Internet]. [S. l.]: StatPearls Publishing, 2024. Citado na p. 30.

REITZ, C. Dyslipidemia and the risk of Alzheimer's disease. **Current atherosclerosis reports**, Springer, v. 15, p. 1–9, 2013. DOI: <https://doi.org/10.1007/s11883-012-0307-3>. Citado na p. 15.

RIZZO, S.; BARTOLINI, M.; CECCARINI, L.; PIAZZI, L.; GOBBI, S.; CAVALLI, A.; RECANATINI, M.; ANDRISANO, V.; RAMPA, A. Targeting Alzheimer's disease: Novel indanone hybrids bearing a pharmacophoric fragment of AP2238. **Bioorganic & medicinal chemistry**, Elsevier, v. 18, n. 5, p. 1749–1760, 2010. DOI: <https://doi.org/10.1016/j.bmc.2010.01.071>. Citado na p. 19.

ROCK, E. M.; PARKER, L. A. Constituents of Cannabis sativa. **Cannabinoids and neuropsychiatric disorders**, Springer, p. 1–13, 2020. DOI: https://doi.org/10.1007/978-3-030-57369-0_1. Citado na p. 22.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986. Citado na p. 45.

RUSSO, E. B. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. **British journal of pharmacology**, Wiley Online Library, v. 163, n. 7, p. 1344–1364, 2011. DOI: <https://doi.org/10.1111/j.1476-5381.2011.01238.x>. Citado na p. 22.

RUSSO, E. B.; BURNETT, A.; HALL, B.; PARKER, K. K. Agonistic properties of cannabidiol at 5-HT_{1a} receptors. **Neurochemical research**, Springer, v. 30, p. 1037–1043, 2005. DOI: <https://doi.org/10.1007/s11064-005-6978-1>. Citado na p. 22.

- SADAMORI, M. **Studies on the new biologically active substances of Tahitian and Tokushima's Plagiochila genus**. 2009. Tese (Doutorado) – Master's Thesis, Tokushima Bunri University, Tokushima, Japan. Citado na p. 23.
- SADEGHI, S.; BUI, A.; FOROOGHI, A.; LU, J.; NGOM, A. Can large language models understand molecules? **BMC bioinformatics**, Springer, v. 25, n. 1, p. 225, 2024. Citado na p. 73.
- SÁNCHEZ, A.; GARCÍA-MERINO, A. Neuroprotective agents: cannabinoids. **Clinical immunology**, Elsevier, v. 142, n. 1, p. 57–67, 2012. DOI: <https://doi.org/10.1016/j.clim.2011.02.010>. Citado na p. 21.
- SCHNEIDER, L. S.; GEFFEN, Y.; RABINOWITZ, J.; THOMAS, R. G.; SCHMIDT, R.; ROPELE, S.; WEINSTOCK, M.; GROUP, L. S. Low-dose ladostigil for mild cognitive impairment: A phase 2 placebo-controlled clinical trial. **Neurology**, Lippincott Williams & Wilkins Hagerstown, MD, v. 93, n. 15, e1474–e1484, 2019. DOI: <https://doi.org/10.1212/wnl.00000000000008239>. Citado na p. 19.
- SCHWARTNER, C.; MICHEL, C.; STETTMAIER, K.; WAGNER, H.; BORS, W. Marchantins and related polyphenols from liverwort: physico-chemical studies of their radical-scavenging properties. **Free Radical Biology and Medicine**, Elsevier, v. 20, n. 2, p. 237–244, 1996. Citado na p. 23.
- SHINTANI, E. Y.; UCHIDA, K. M. Donepezil: an anticholinesterase inhibitor for Alzheimer's disease. **American journal of health-system pharmacy**, Oxford University Press, v. 54, n. 24, p. 2805–2810, 1997. Citado na p. 17.
- SHIRBHATE, E.; PATEL, V. K.; TIWARI, P.; KORE, R.; VEERASAMY, R.; MISHRA, A.; RAJAK, H. Combination Therapy for the Treatment of Alzheimer's Disease: Recent Progress and Future Prospects. **Current Topics in Medicinal Chemistry**, Bentham Science Publishers, v. 22, n. 22, p. 1849–1867, 2022. Citado na p. 19.
- SILVA, M. A.; KIAMETIS, A. S.; TREPTOW, W. Donepezil inhibits acetylcholinesterase via multiple binding modes at room temperature. **Journal of Chemical Information and Modeling**, ACS Publications, v. 60, n. 7, p. 3463–3471, 2020. Citado na p. 65.
- SIMONS, R.; GRUPPEN, H.; BOVEE, T. F.; VERBRUGGEN, M. A.; VINCKEN, J.-P. Prenylated isoflavonoids from plants as selective estrogen receptor modulators (phytoSERMs). **Food & function**, Royal Society of Chemistry, v. 3, n. 8, p. 810–827, 2012. Citado na p. 25.
- SINGH, K.; BHUSHAN, B.; CHANCHAL, D. K.; SHARMA, S. K.; RANI, K.; YADAV, M. K.; PORWAL, P.; KUMAR, S.; SHARMA, A.; VIRMANI, T. *et al.* Emerging therapeutic potential of cannabidiol (CBD) in neurological disorders: a comprehensive review. **Behavioural neurology**, Wiley Online Library, v. 2023, n. 1, p. 8825358, 2023. DOI: <https://doi.org/10.1155/2023/8825358>. Citado na p. 22.
- SMITH, M. A.; HARRIS, P. L.; SAYRE, L. M.; PERRY, G. Iron accumulation in Alzheimer disease is a source of redox-generated free radicals. **Proceedings of the National Academy of Sciences**, The National Academy of Sciences of the USA, v. 94, n. 18, p. 9866–9868, 1997. DOI: <https://doi.org/10.1073/pnas.94.18.9866>. Citado na p. 14.
- SOFFER, S.; BEN-COHEN, A.; SHIMON, O.; AMITAI, M. M.; GREENSPAN, H.; KLANG, E. Convolutional neural networks for radiologic images: a radiologist's guide. **Radiology**, Radiological Society of North America, v. 290, n. 3, p. 590–606, 2019. Citado na p. 47.
- SØNDERGAARD, C. R.; OLSSON, M. H. M.; ROSTKOWSKI, M.; JENSEN, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. **Journal of Chemical Theory and Computation**, American Chemical Society (ACS), v. 7, n. 7, p. 2284–2295, jun. 2011. ISSN 1549-9626. DOI: 10.1021/ct200133y. Disponível em: <http://dx.doi.org/10.1021/ct200133y>. Citado na p. 48.

- STANCIU, G. D.; LUCA, A.; RUSU, R. N.; BILD, V.; BESCHEA CHIRIAC, S. I.; SOLCAN, C.; BILD, W.; ABABEI, D. C. Alzheimer's disease pharmacotherapy in relation to cholinergic system involvement. **Biomolecules**, MDPI, v. 10, n. 1, p. 40, 2019. DOI: <https://doi.org/10.3390/biom10010040>. Citado na p. 14.
- STEPHENS, P. J.; DEVLIN, F. J.; CHABALOWSKI, C. F.; FRISCH, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. **The Journal of physical chemistry**, ACS Publications, v. 98, n. 45, p. 11623–11627, 1994. Citado na p. 49.
- STOCK, L.; HOSOUME, J.; TREPTOW, W. Concentration-dependent binding of small ligands to multiple saturable sites in membrane proteins. **Scientific Reports**, Nature Publishing Group UK London, v. 7, n. 1, p. 5734, 2017. Citado na p. 65.
- STONE, N. L.; MURPHY, A. J.; ENGLAND, T. J.; O'SULLIVAN, S. E. A systematic review of minor phytocannabinoids with promising neuroprotective potential. **British Journal of Pharmacology**, Wiley Online Library, v. 177, n. 19, p. 4330–4352, 2020. Citado na p. 21.
- SUMMERS, W. K. **Administration of monoamine acridines in cholinergic neuronal deficit states**. [S. l.]: Google Patents, mar. 1989. US Patent 4,816,456. Citado na p. 17.
- SUMMERS, W. K.; VIESSELMAN, J. O.; MARSH, G. M.; CANDELORA, K. Use of THA in treatment of Alzheimer-like dementia: pilot study in twelve patients. **Biological Psychiatry**, v. 16, n. 2, p. 145–153, 1981. Citado na p. 17.
- TADMOR, Y.; TOLHURST, D. J. Discrimination of changes in the second-order statistics of natural and synthetic images. **Vision research**, Elsevier, v. 34, n. 4, p. 541–554, 1994. Citado na p. 39.
- TAI, H.-C.; WANG, B. Y.; SERRANO-POZO, A.; FROSC, M. P.; SPIRES-JONES, T. L.; HYMAN, B. T. Frequent and symmetric deposition of misfolded tau oligomers within presynaptic and postsynaptic terminals in Alzheimer's disease. **Acta neuropathologica communications**, Springer, v. 2, p. 1–14, 2014. DOI: <https://doi.org/10.1186/s40478-014-0146-2>. Citado na p. 14.
- TAKADA-TAKATORI, Y.; KUME, T.; OHGI, Y.; IZUMI, Y.; NIIDOME, T.; FUJII, T.; SUGIMOTO, H.; AKAIKE, A. Mechanism of neuroprotection by donepezil pretreatment in rat cortical neurons chronically treated with donepezil. **Journal of neuroscience research**, Wiley Online Library, v. 86, n. 16, p. 3575–3583, 2008. Citado na p. 17.
- TAKASHIMA, A. Hyperphosphorylated tau is a cause of neuronal dysfunction in tauopathy. **Journal of Alzheimer's disease**, SAGE Publications Sage UK: London, England, v. 14, n. 4, p. 371–375, 2008. DOI: <https://doi.org/10.3233/jad-2008-14403>. Citado na p. 14.
- TAKEDA, M.; TAGAMI, S. A history of antidementic drug development in Japan. **Taiwanese Journal of Psychiatry**, Medknow, v. 34, n. 4, p. 152–161, 2020. Citado na p. 17.
- TEAM, T. pandas development. **pandas-dev/pandas: Pandas**. [S. l.]: Zenodo, fev. 2020. DOI: 10.5281/zenodo.3509134. Disponível em: <https://doi.org/10.5281/zenodo.3509134>. Citado na p. 49.
- TING, F. F.; TAN, Y. J.; SIM, K. S. Convolutional neural network improvement for breast cancer classification. **Expert Systems with Applications**, Elsevier, v. 120, p. 103–115, 2019. Citado na p. 47.
- TITAR, R. R.; RAMANATHAN, M. Variational autoencoders for generative modeling of drug dosing determinants in renal, hepatic, metabolic, and cardiac disease states. **Clinical and Translational Science**, Wiley Online Library, v. 17, n. 7, e13872, 2024. Citado na p. 46.
- TONIUTTI, M. A.; FORNASERO, L. V.; ALBICORO, F. J.; MARTINI, M. C.; DRAGHI, W.; ALVAREZ, F.; LAGARES, A.; PENSIERO, J. F.; DEL PAPA, M. F. Nitrogen-fixing rhizobial strains

isolated from *Desmodium incanum* DC in Argentina: Phylogeny, biodiversity and symbiotic ability. **Systematic and Applied Microbiology**, Elsevier, v. 40, n. 5, p. 297–307, 2017. Citado na p. 25.

TORNG, W.; ALTMAN, R. B. Graph convolutional neural networks for predicting drug-target interactions. **Journal of chemical information and modeling**, ACS Publications, v. 59, n. 10, p. 4131–4149, 2019. Citado na p. 47.

TOYOTA, M.; KINUGAWA, T.; ASAKAWA, Y. Bibenzyl cannabinoid and bisbibenzyl derivative from the liverwort *Radula perrottetii*. **Phytochemistry**, Elsevier, v. 37, n. 3, p. 859–862, 1994. Citado na p. 23.

TOYOTA, M.; SHIMAMURA, T.; ISHII, H.; RENNER, M.; BRAGGINS, J.; ASAKAWA, Y. New bibenzyl cannabinoid from the New Zealand liverwort *Radula marginata*. **Chemical and pharmaceutical bulletin**, The Pharmaceutical Society of Japan, v. 50, n. 10, p. 1390–1392, 2002. Citado na p. 23.

TREFETHEN, L. N.; BAU, D. **Numerical linear algebra**. [S. l.]: SIAM, 2022. Citado na p. 43.

TREJO-LOPEZ, J. A.; YACHNIS, A. T.; PROKOP, S. Neuropathology of Alzheimer's disease. **Neurotherapeutics**, Elsevier, v. 19, n. 1, p. 173–185, 2022. DOI: <https://doi.org/10.1007/s13311-021-01146-y>. Citado na p. 14.

TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of computational chemistry**, Wiley Online Library, v. 31, n. 2, p. 455–461, 2010. Citado na p. 30.

TURING, A. M. Equivalence of left and right almost periodicity. **Journal of the London Mathematical Society**, Oxford University Press, v. 1, n. 4, p. 284–285, 1935. Citado na p. 44.

TYRAKIS, P.; AGRIDI, C.; KOURTI, M. A Comprehensive Exploration of the Multifaceted Neuroprotective Role of Cannabinoids in Alzheimer's Disease across a Decade of Research. **International Journal of Molecular Sciences**, MDPI, v. 25, n. 16, p. 8630, 2024. DOI: <https://doi.org/10.3390/ijms25168630>. Citado na p. 21.

TZIORAS, M.; MCGEACHAN, R. I.; DURRANT, C. S.; SPIRES-JONES, T. L. Synaptic degeneration in Alzheimer disease. **Nature Reviews Neurology**, Nature Publishing Group UK London, v. 19, n. 1, p. 19–38, 2023. Citado na p. 13.

UZUNOVA, H.; SCHULTZ, S.; HANDELS, H.; EHRHARDT, J. Unsupervised pathology detection in medical images using conditional variational autoencoders. **International journal of computer assisted radiology and surgery**, Springer, v. 14, p. 451–461, 2019. Citado na p. 46.

VANI, N.; SOWMYA, A.; JAYAMMA, N. Brain tumor classification using support vector machine. **International Research Journal of Engineering and Technology (IRJET)**, v. 4, n. 7, p. 792–796, 2017. Citado na p. 32.

VAPNIK, V.; CHERVONENKIS, A. Y. A class of algorithms for pattern recognition learning. **Avtomat. i Telemekh**, v. 25, n. 6, p. 937–945, 1964. Citado na p. 32.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017. Citado na p. 73.

VECCHIO, I.; SORRENTINO, L.; PAOLETTI, A.; MARRA, R.; ARBITRIO, M. The state of the art on acetylcholinesterase inhibitors in the treatment of Alzheimer's disease. **Journal of Central nervous system disease**, SAGE Publications Sage UK: London, England, v. 13, p. 11795735211029113, 2021. DOI: <https://doi.org/10.1177/11795735211029113>. Citado na p. 17.

- VERHULST, P.-F. Sur la loi d'accroissement de la population (Deuxième mémoire). **Mémoires de l'Académie royale de Belgique**, Persée-Portail des revues scientifiques en SHS, v. 20, n. 1, p. 1–32, 1847. Citado na p. 44.
- VOSKO, S. H.; WILK, L.; NUSAIR, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. **Canadian Journal of physics**, NRC Research Press Ottawa, Canada, v. 58, n. 8, p. 1200–1211, 1980. Citado na p. 49.
- WALLACH, I.; DZAMBA, M.; HEIFETS, A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. **arXiv preprint arXiv:1510.02855**, 2015. Citado na p. 47.
- WALSH, K. B.; MCKINNEY, A. E.; HOLMES, A. E. Minor cannabinoids: Biosynthesis, molecular pharmacology and potential therapeutic uses. **Frontiers in pharmacology**, Frontiers Media SA, v. 12, p. 777804, 2021. DOI: <https://doi.org/10.3389/fphar.2021.777804>. Citado na p. 22.
- WARSHEL, A. Energetics of enzyme catalysis. **Proceedings of the National Academy of Sciences**, v. 75, n. 11, p. 5250–5254, 1978. DOI: <https://doi.org/10.1073/pnas.75.11.5250>. Citado na p. 29.
- WATKINS, D. S. **Fundamentals of matrix computations**. [S. l.]: John Wiley & Sons, 2004. Citado na p. 41.
- WATKINS, P. B.; ZIMMERMAN, H. J.; KNAPP, M. J.; GRACON, S. I.; LEWIS, K. W. Hepatotoxic effects of tacrine administration in patients with Alzheimer's disease. **Jama**, American Medical Association, v. 271, n. 13, p. 992–998, 1994. DOI: <https://doi.org/10.1001/jama.1994.03510370044030>. Citado na p. 17.
- WEINREB, O.; AMIT, T.; BAR-AM, O.; YODIM, M. B. A novel anti-Alzheimer's disease drug, ladostigil: neuroprotective, multimodal brain-selective monoamine oxidase and cholinesterase inhibitor. **International Review of Neurobiology**, Elsevier, v. 100, p. 191–215, 2011. DOI: <https://doi.org/10.1016/b978-0-12-386467-3.00010-8>. Citado na p. 19.
- WEINREB, O.; MANDEL, S.; BAR-AM, O.; YOGEV-FALACH, M.; AVRAMOVICH-TIROSH, Y.; AMIT, T.; YODIM, M. B. Multifunctional neuroprotective derivatives of rasagiline as anti-Alzheimer's disease drugs. **Neurotherapeutics**, Elsevier, v. 6, n. 1, p. 163–174, 2009. Citado na p. 19.
- WEINSTOCK, M.; LUQUES, L.; BEJAR, C.; SHOHAM, S. Ladostigil, a novel multifunctional drug for the treatment of dementia co-morbid with depression. **Parkinson's Disease and Related Disorders**, Springer, p. 443–446, 2006. Citado na p. 19.
- WHO, W. H. O. **Ageing and health**. [S. l.: s. n.], 2022. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>. [Accessed 09-05-2024]. Citado na p. 13.
- WILKINSON, D.; MURRAY, J.; GROUP, G. R. Galantamine: a randomized, double-blind, dose comparison in patients with Alzheimer's disease. **International journal of geriatric psychiatry**, Wiley Online Library, v. 16, n. 9, p. 852–857, 2001. DOI: <https://doi.org/10.1002/gps.409>. Citado na p. 18.
- WILKINSON, D. G. The pharmacology of donepezil: a new treatment for Alzheimer's disease. **Expert opinion on pharmacotherapy**, Taylor & Francis, v. 1, n. 1, p. 121–135, 1999. Citado na p. 17.
- WU, T.; LIN, D.; CHENG, Y.; JIANG, S.; RIAZ, M. W.; FU, N.; MOU, C.; YE, M.; ZHENG, Y. Amyloid cascade hypothesis for the treatment of alzheimer's disease: progress and challenges. **Aging and Disease**, JKL International LLC, v. 13, n. 6, p. 1745, 2022. Citado na p. 19.

WU, W.; WANG, X.; XU, Y.; MA, C.; QIAN, X.; QIU, W. Neuropathological comorbidity, genetics and cognition in a Chinese community-based autopsy cohort. **Brain**, Oxford University Press, awaf039, 2025. DOI: <https://doi.org/10.1093/brain/awaf039>. Citado na p. 15.

YADAV, S. S.; JADHAV, S. M. Deep convolutional neural network based medical image classification for disease diagnosis. **Journal of Big data**, Springer, v. 6, n. 1, p. 1–18, 2019. Citado na p. 47.

YAMADA, K. A revision of Asian taxa of Radula, Hepaticae. **The Journal of the Hattori Botanical Laboratory**, Hattori Botanical Laboratory, v. 45, p. 201–322, 1979. Citado na p. 22.

YANG, G.; WANG, Y.; TIAN, J.; LIU, J.-P. Huperzine A for Alzheimer’s disease: a systematic review and meta-analysis of randomized clinical trials. **PloS one**, Public Library of Science San Francisco, USA, v. 8, n. 9, e74916, 2013. Citado na p. 18.

ZANETTINI, C.; PANLILIO, L. V.; ALICZKI, M.; GOLDBERG, S. R.; HALLER, J.; YASAR, S. Effects of endocannabinoid system modulation on cognitive and emotional behavior. **Frontiers in behavioral neuroscience**, Frontiers Research Foundation, v. 5, p. 57, 2011. DOI: <https://doi.org/10.3389/fnbeh.2011.00057>. Citado na p. 21.

ZHANG, C.; YAN, Q.; ZHU, Y.; LI, X.; SUN, J.; ZHANG, Y. Attention-based network for low-light image enhancement. *In*: IEEE. 2020 IEEE International Conference on Multimedia and Expo (ICME). [S. l.: s. n.], 2020. p. 1–6. Citado na p. 46.

ZHANG, J.; LIN, C.; JIN, S.; WANG, H.; WANG, Y.; DU, X.; HUTCHINSON, M. R.; ZHAO, H.; FANG, L.; WANG, X. The pharmacology and therapeutic role of cannabidiol in diabetes. *In*: WILEY ONLINE LIBRARY, 5. EXPLORATION. [S. l.: s. n.], 2023. v. 3, p. 20230047. DOI: <https://doi.org/10.1002/exp.20230047>. Citado na p. 22.

ZIADA, S.; DIHARCE, J.; RAIMBAUD, E.; ACI-SÈCHE, S.; DUCROT, P.; BONNET, P. Estimation of drug-target residence time by targeted molecular dynamics simulations. **Journal of chemical information and modeling**, ACS Publications, v. 62, n. 22, p. 5536–5549, 2022. Citado na p. 30.

ZUBAIR, A. R.; ALO, O. A. Grey level co-occurrence matrix (GLCM) based second order statistics for image texture analysis. **arXiv preprint arXiv:2403.04038**, 2024. Citado na p. 39.

Capítulo 7

APÊNDICES

7.1 Tabelas suplementares

Figura 7.1: Interações não covalentes entre as 32 moléculas estudadas e a cavidade GORGE da enzima hAChE.

GORGE	H447	W86	Y337	P338	Y341	Y124	W286	P295	S293	D74	SCORE
<i>DNP</i>	3,60	4,26	3,75	3,88	4,83	5,50	3,55	1,89	3,05	5,00	-11,60
<i>PRT</i>	6,00	3,86	4,40	5,08	4,67	6,00	3,62	6,00	2,73	3,36	-10,32
<i>DMDD</i>	4,78	3,71	3,44	4,44	5,25	6,00	4,51	6,00	2,90	4,55	-9,70
<i>MCRA</i>	4,30	5,30	3,58	5,34	4,57	5,47	4,60	6,00	6,00	2,63	-9,48
<i>DMDE</i>	6,00	4,84	3,82	5,27	4,84	3,38	4,92	6,00	6,00	4,24	-9,31
<i>MCDC</i>	4,96	3,50	3,24	5,01	5,79	5,32	4,45	6,00	6,00	4,68	-8,82
<i>OCMNB</i>	4,11	4,62	5,02	5,35	4,80	5,65	4,55	6,00	6,00	3,30	-8,72
<i>D9CT</i>	4,41	4,13	6,45	4,92	5,05	3,52	4,54	5,93	6,00	6,00	-8,61
<i>HHC</i>	6,00	3,89	3,74	3,28	3,61	2,76	3,00	6,00	6,00	2,27	-8,48
<i>HD11</i>	4,87	5,01	3,95	4,20	3,99	3,47	3,38	2,63	3,54	4,21	-8,39
<i>ARAC1</i>	3,65	3,54	4,18	4,68	4,76	5,90	4,86	3,10	2,90	6,00	-8,09
<i>CNBRPL</i>	5,04	4,89	5,77	5,93	2,78	4,31	4,89	2,27	6,02	7,19	-8,00
<i>DDAA</i>	4,00	4,50	3,60	3,79	4,28	2,50	3,56	3,61	3,15	4,54	-7,82
<i>ARAC2</i>	4,45	4,93	2,94	3,39	5,35	2,14	4,74	3,03	2,18	4,99	-7,69
<i>GLDC</i>	3,40	5,00	3,42	3,71	2,16	3,51	4,39	1,76	3,54	3,15	-7,63
<i>CNBA</i>	5,53	6,00	4,62	4,37	4,17	3,47	3,84	2,33	3,89	4,31	-7,27
<i>THC</i>	4,96	4,83	5,85	6,03	6,67	6,55	5,29	6,00	6,00	6,00	-7,26
<i>R6E</i>	5,48	4,43	4,97	3,97	4,10	2,47	3,25	2,94	4,74	6,71	-7,15
<i>DCBA</i>	5,46	6,74	5,90	4,45	4,67	4,81	3,65	6,43	3,89	4,85	-7,14
<i>RDLJ</i>	4,36	3,39	3,49	5,56	5,50	4,36	3,72	5,11	5,24	5,89	-7,12
<i>RDLI</i>	3,70	4,11	3,23	4,50	4,88	2,80	4,38	2,053	5,19	4,05	-7,10
<i>RDLA</i>	4,03	4,86	3,43	3,92	3,79	4,81	3,64	5,49	6,76	5,20	-6,95

Figura 7.2: Interações não covalentes entre as 32 moléculas estudadas e a cavidade GORGE da enzima hAChE.

<i>CNDB</i>	4,85	6,14	3,53	4,63	5,05	5,48	3,11	6,00	3,87	5,60	-6,77
<i>RDL59</i>	3,94	6,32	6,59	3,16	4,88	3,70	3,63	2,52	5,38	3,21	-6,66
<i>THCP</i>	5,24	3,55	4,90	5,69	4,24	6,41	4,27	2,42	3,27	5,35	-6,48
<i>RDLK</i>	3,84	3,72	4,54	3,94	4,39	5,051	4,16	1,96	4,90	4,313	-6,36
<i>CMG</i>	4,72	4,14	3,17	3,67	4,86	4,53	3,32	1,83	3,36	4,12	-6,33
<i>ITHC</i>	5,68	5,24	4,86	3,45	3,81	5,94	4,21	2,31	4,50	2,05	-6,18
<i>RDLH</i>	3,12	6,10	4,15	4,97	4,40	4,17	3,58	2,31	6,33	3,60	-5,92
<i>TCNT</i>	4,11	4,03	4,92	3,30	3,46	3,69	3,19	4,15	5,92	3,99	-5,07
<i>THCVM2</i>	5,59	7,64	5,078	4,26	4,60	4,85	5,77	2,88	1,98	5,07	-4,99
<i>LVD</i>	10,06	8,23	7,78	5,73	4,81	1,85	6,27	3,32	2,07	2,02	-4,62
van der Waals	Ligaçãode Hidrogênio	Lig. H. não convencional	LH pi-doadora								
Pi-alquil	Pi-pi T-Stacked	Pi-pi Stacked	Pi-sigma								
Cátion-pi	LUMO +1	LUMO	HOMO								

Tabela 7.1: Descritores de estrutura eletrônica calculados para as moléculas do trabalho.

Molécula	HOMO-4 (eV)	HOMO-3 (eV)	HOMO-2 (eV)	HOMO-1 (eV)	HOMO (eV)	LUMO (eV)	LUMO+1 (eV)	LUMO+2 (eV)	LUMO+3 (eV)	LUMO+4 (eV)	GAP	CDD ⁺	CDD ⁻
DNP	-7.20	-7.19	-7.15	-6.98	-6.15	-1.74	-0.89	-0.59	-0.37	0.09	4.42	0.146	-0.084
LVD	-9.05	-8.30	-7.06	-6.76	-6.17	-0.88	-0.48	-0.08	-0.02	0.23	5.28	0.155	-0.086
CMG	-7.74	-7.63	-6.64	-6.23	-6.11	-0.22	-0.01	0.06	0.20	0.35	5.89	0.103	-0.105
CNBRPL	-7.02	-7.63	-7.45	-6.27	-6.00	-0.24	0.02	0.04	0.35	0.46	5.76	0.128	-0.097
GLDC	-7.94	-7.00	-6.80	-6.28	-6.13	-0.54	-0.42	-0.19	-0.02	0.05	5.59	0.093	-0.090
OMCNB	-7.69	-7.54	-6.73	-6.44	-6.25	-1.23	-0.40	-0.16	0.15	0.38	4.68	0.017	-0.033
PRT	-7.01	-6.80	-6.35	-6.46	-6.00	-0.50	-0.42	-0.10	0.02	0.06	5.50	0.098	-0.087
THC	-7.94	-7.87	-6.46	-5.95	-5.95	-0.19	0.06	0.10	0.37	0.41	5.75	0.133	-0.083
THCP	-7.96	-7.87	-6.45	-6.18	-5.97	-0.21	0.04	0.10	0.36	0.42	5.76	0.126	-0.083
THCVM2	-8.08	-8.03	-7.29	-6.32	-6.03	-1.54	-0.24	0.00	0.03	0.35	4.49	0.096	-0.098
CNBA	-8.13	-7.75	-6.84	-6.38	-6.06	-1.34	-1.02	-0.54	0.02	0.12	4.72	0.020	-0.035
HHC	-6.73	-6.44	-6.37	-6.28	-6.19	-1.12	-0.52	-0.31	-0.12	0.00	5.07	0.080	-0.046
CNBD	-7.26	-7.09	-6.54	-6.35	-6.16	-0.82	-0.53	-0.15	0.02	0.04	5.34	0.068	-0.072
ARAC1	-7.56	-6.76	-6.31	-6.23	-5.43	-1.75	-0.51	-0.17	-0.03	0.04	3.67	0.021	-0.019
ARAC2	-7.22	-6.79	-6.47	-6.20	-5.52	-1.60	-0.56	-0.03	0.00	0.15	3.92	0.018	-0.017
MCDC	-7.03	-6.77	-6.70	-6.22	-5.78	-1.55	-0.31	-0.10	-0.03	0.11	4.23	0.014	-0.015
MCDD	-7.40	-6.98	-6.75	-6.15	-5.78	-1.54	-0.30	-0.13	0.08	0.16	4.23	0.016	-0.022
RDLA	-7.02	-6.84	-6.61	-6.49	-6.24	-0.55	-0.42	-0.25	-0.07	0.03	5.69	0.130	-0.050
RDLK	-7.79	-7.21	-6.75	-6.14	-5.54	-1.59	-0.56	-0.07	0.08	0.16	3.94	0.041	-0.025
RDLI	-7.69	-7.11	-6.79	-6.24	-5.63	-1.72	-0.48	-0.15	-0.03	0.14	3.91	0.017	-0.022
RDLH	-7.03	-6.92	-6.78	-6.63	-6.43	-1.72	-0.45	-0.39	-0.21	-0.02	4.71	0.081	-0.036
RDLJ	-7.62	-7.11	-6.78	-6.23	-5.65	-1.75	-0.48	-0.09	0.03	0.19	3.90	0.024	-0.026
RDL59	-7.01	-6.97	-6.75	-6.69	-6.59	-1.49	-0.70	-0.42	-0.34	-0.01	5.10	0.062	-0.054
DCBA	-7.01	-6.81	-6.65	-6.21	-6.09	-0.52	-0.42	-0.17	0.06	0.14	5.57	0.033	-0.068
DMDD	-7.16	-6.68	-6.37	-6.24	-6.09	-1.87	-0.48	-0.15	-0.08	0.04	4.22	0.099	-0.068
DMDE	-6.68	-6.46	-6.42	-6.25	-6.07	-1.86	-0.39	-0.14	-0.06	-0.01	4.20	0.097	-0.070
D9CT	-7.98	-7.85	-6.44	-6.21	-5.97	-0.17	0.03	0.09	0.31	0.47	5.80	0.138	-0.087
TCNT	-7.92	-7.76	-7.43	-6.16	-5.78	-0.95	-0.04	0.04	0.35	0.43	4.83	0.118	-0.033
HD11	-8.02	-7.61	-6.92	-6.22	-5.99	-0.31	-0.13	0.06	0.20	0.40	5.68	0.060	-0.107
DDAA	-7.02	-6.81	-6.57	-6.22	-6.13	-0.52	-0.42	-0.14	0.01	0.05	5.61	0.030	-0.068
MCRA	-7.64	-7.11	-6.88	-6.08	-5.69	-1.70	-0.48	-0.03	0.03	0.12	3.99	0.024	-0.030
R6E	-8.08	-7.56	-6.39	-6.36	-6.13	-0.26	0.02	0.04	0.32	0.37	5.87	0.154	-0.062

Tabela 7.2: Descritores ADMET e valores de Score-ADT em kcal/mol

Molécula	Comprimento (Å)	Volume (cm ³)	MM	Átomos aromáticos	Fração Csp ³	Ligação-H acceptor	Ligação-H doador	RM	TPSA	Log P	Permeabilidade BHE	Violações de Lipinski	Drugscore	Score (ADT)
DNP	16.28	310.53	380.50	12.00	0.46	3.00	1.00	116.27	47.56	3.11	1.00	0.00	0.63	-11.6
LVD	8.60	161.31	197.19	6.00	0.22	5.00	4.00	49.55	103.78	-0.86	2.00	0.00	0.49	-4.62
CMG	17.24	337.47	350.49	6.00	0.62	4.00	4.00	104.80	80.92	4.13	2.00	0.00	0.30	-6.33
CNBRPL	15.95	259.95	348.48	6.00	0.71	4.00	3.00	100.74	69.92	3.65	1.00	0.00	0.32	-8
GLDC	16.27	248.74	298.38	12.00	0.26	3.00	3.00	89.63	60.69	3.62	1.00	0.00	0.37	-7.63
OMCNB	15.27	213.79	324.46	12.00	0.45	2.00	0.00	101.56	18.46	5.58	2.00	-1.25	0.18	-8.72
PRT	15.86	317.08	346.46	12.00	0.33	2.00	1.00	109.09	29.46	5.43	2.00	-1.25	0.12	-10.62
THC	15.37	223.41	314.46	6.00	0.62	2.00	1.00	97.91	29.46	5.33	1.00	-1.25	0.23	-7.26
THCP	17.88	221.94	342.51	6.00	0.65	2.00	1.00	107.52	29.46	5.92	2.00	-1.25	0.18	-6.48
THCVM2	13.17	257.97	316.39	6.00	0.53	4.00	2.00	90.07	66.76	3.72	1.00	0.00	0.37	-4.99
CNBA	15.26	319.58	354.44	12.00	0.41	4.00	2.00	104.05	66.76	4.84	2.00	0.00	0.20	-7.27
HHC	20.17	383.10	410.50	12.00	0.32	5.00	4.00	121.01	97.99	5.11	2.00	0.00	0.24	-8.48
CNBD	15.58	246.28	310.43	12.00	0.33	2.00	2.00	90.98	40.46	5.52	1.00	-1.25	0.22	-6.77
ARAC1	15.28	225.19	312.36	12.00	0.16	4.00	4.00	94.42	80.92	3.61	2.00	0.00	0.37	-8.09
ARAC2	16.10	251.56	298.38	12.00	0.26	3.00	3.00	90.49	60.69	4.00	1.00	0.00	0.41	-7.69
MCDC	15.14	313.87	364.48	15.00	0.33	3.00	2.00	111.34	53.60	5.38	2.00	0.00	0.10	-8.82
MCDD	15.97	254.45	378.46	15.00	0.42	4.00	2.00	110.56	62.83	4.37	2.00	0.00	0.20	-9.99
RDLA	14.60	226.64	278.35	12.00	0.16	2.00	1.00	87.13	29.46	4.01	1.00	0.00	0.23	-6.95
RDLK	15.11	234.33	312.36	12.00	0.26	4.00	3.00	90.82	69.92	3.00	1.00	0.00	0.31	-6.36
RDLI	14.62	227.40	278.35	12.00	0.26	2.00	1.00	85.23	29.46	4.01	1.00	0.00	0.25	-7.1
RDLH	14.21	225.95	322.35	12.00	0.15	4.00	2.00	94.09	66.76	3.66	1.00	0.00	0.23	-5.92
RDLJ	14.55	230.98	292.37	12.00	0.30	2.00	0.00	94.70	18.46	4.43	1.00	0.00	0.25	-7.12
RDL59	14.14	203.76	322.35	12.00	0.15	4.00	2.00	89.09	66.76	3.52	1.00	0.00	0.23	-6.66
DCBA	15.38	257.90	296.40	12.00	0.30	2.00	1.00	92.94	29.46	4.79	1.00	-1.25	0.29	-7.14
DMDD	16.14	357.31	436.50	12.00	0.50	6.00	3.00	119.89	96.22	4.05	2.00	0.00	0.16	-9.7
DMDE	16.12	343.48	436.50	12.00	0.42	6.00	3.00	121.79	96.22	4.17	2.00	0.00	0.14	-9.31
D9CT	15.35	262.35	314.46	6.00	0.62	2.00	1.00	97.91	29.46	5.28	1.00	-1.25	0.23	-8.61
TCNT	15.03	255.40	332.43	6.00	0.60	4.00	3.00	96.22	69.92	3.26	1.00	0.00	0.14	-5.07
HD11	14.76	278.61	330.46	6.00	0.62	3.00	2.00	99.07	49.69	4.55	1.00	0.00	0.27	-8.39
DDAA	15.63	219.66	282.38	12.00	0.26	2.00	2.00	88.47	40.46	4.42	1.00	0.00	0.36	-7.82
MCRA	15.92	308.93	348.48	12.00	0.42	2.00	1.00	109.56	29.46	5.38	2.00	-1.25	0.11	-9.99
R6E	13.62	246.50	332.48	6.00	0.62	3.00	2.00	101.45	52.99	5.02	1.00	0.00	0.21	-7.15

Tabela 7.3: Matriz de similaridade usando similaridade de Tanimoto, S_T

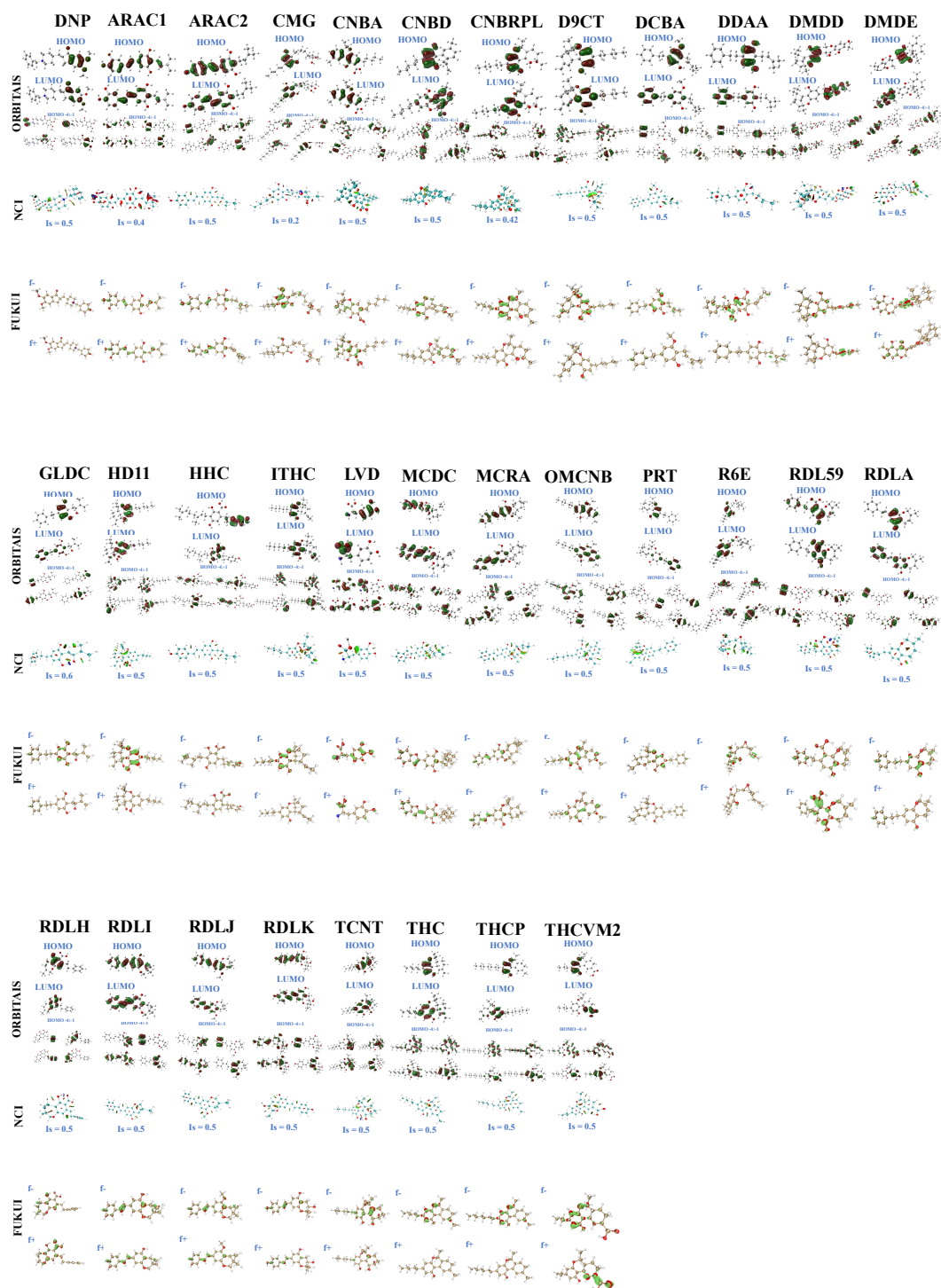
MOLÉCULA	DNP	LVD	CMG	CNBRPL	GLDC	OMCNB	PRT	THC	THCP	THCVM2	CNBA	HHC	CNBD	ARAC1	ARAC2	MCDC	MCDD	RDLA	RDLK	RDLI	RDLH	RDLJ	RDL59	DCBA	DMDD	DMDE	D9CT	TCNT	HD11	DDAA	MCRA	R6E																																																																																																																																																																																																																																																																																																	
DNP	1.0	0.121	0.071	0.106	0.159	0.105	0.134	0.106	0.105	0.104	0.086	0.085	0.12	0.101	0.164	0.146	0.145	0.122	0.219	0.136	0.235	0.109	0.109	0.106	0.119	0.103	0.167	0.134	0.167	0.134	0.165																																																																																																																																																																																																																																																																																																		
LVD	0.121	1.0	0.186	0.138	0.304	0.108	0.123	0.138	0.136	0.19	0.152	0.23	0.204	0.244	0.188	0.123	0.099	0.204	0.138	0.136	0.143	0.078	0.161	0.231	0.093	0.093	0.138	0.138	0.134	0.267	0.123	0.169																																																																																																																																																																																																																																																																																																	
CMG	0.071	0.186	1.0	0.304	0.288	0.222	0.099	0.268	0.264	0.182	0.179	0.3	0.328	0.143	0.281	0.112	0.08	0.174	0.084	0.107	0.086	0.062	0.1	0.311	0.089	0.077	0.268	0.268	0.278	0.377	0.099	0.574																																																																																																																																																																																																																																																																																																	
CNBRPL	0.106	0.138	0.304	1.0	0.176	0.114	0.31	0.593	0.583	0.462	0.247	0.159	0.288	0.101	0.149	0.177	0.289	0.232	0.16	0.298	0.108	0.141	0.108	0.2	0.214	0.229	0.593	0.446	0.684	0.222	0.31	0.413																																																																																																																																																																																																																																																																																																	
GLDC	0.159	0.204	0.288	0.176	1.0	0.147	0.162	0.178	0.174	0.191	0.139	0.286	0.291	0.17	0.255	0.197	0.135	0.365	0.143	0.177	0.147	0.119	0.164	0.458	0.1	0.1	0.176	0.176	0.138	0.587	0.162	0.25																																																																																																																																																																																																																																																																																																	
OMCNB	0.105	0.108	0.222	0.314	0.147	1.0	0.152	0.333	0.329	0.224	0.567	0.107	0.407	0.104	0.103	0.096	0.129	0.203	0.122	0.133	0.111	0.132	0.098	0.242	0.163	0.163	0.333	0.284	0.396	0.154	0.138	0.279																																																																																																																																																																																																																																																																																																	
PRT	0.134	0.123	0.099	0.31	0.162	0.152	1.0	0.576	0.567	0.469	0.173	0.119	0.135	0.25	0.267	0.211	0.229	0.235	0.348	0.599	0.3	0.275	0.2	0.169	0.202	0.403	0.576	0.298	0.328	0.188	0.614	0.173																																																																																																																																																																																																																																																																																																	
THC	0.106	0.138	0.268	0.593	0.176	0.333	0.576	1.0	0.979	0.638	0.263	0.118	0.328	0.134	0.149	0.177	0.273	0.269	0.19	0.243	0.15	0.171	0.15	0.167	0.2	0.397	1.0	0.446	0.627	0.185	0.229	0.369																																																																																																																																																																																																																																																																																																	
THCP	0.105	0.136	0.264	0.583	0.174	0.329	0.567	0.979	1.0	0.627	0.2	0.118	0.328	0.134	0.149	0.177	0.273	0.269	0.19	0.243	0.15	0.171	0.15	0.167	0.2	0.397	1.0	0.446	0.627	0.185	0.229	0.369																																																																																																																																																																																																																																																																																																	
THCVM2	0.104	0.135	0.262	0.582	0.173	0.328	0.567	0.979	0.627	1.0	0.2	0.118	0.328	0.134	0.149	0.177	0.273	0.269	0.19	0.243	0.15	0.171	0.15	0.167	0.2	0.397	1.0	0.446	0.627	0.185	0.229	0.369																																																																																																																																																																																																																																																																																																	
CNBA	0.083	0.112	0.142	0.162	0.191	0.124	0.169	0.163	0.167	0.1	0.244	0.171	0.211	0.116	0.13	0.175	0.209	0.146	0.234	0.239	0.162	0.169	0.182	0.181	0.188	0.321	0.638	0.38	0.402	0.2	0.324	0.088																																																																																																																																																																																																																																																																																																	
HHC	0.083	0.112	0.142	0.162	0.191	0.124	0.169	0.163	0.167	0.1	0.244	0.171	0.211	0.116	0.13	0.175	0.209	0.146	0.234	0.239	0.162	0.169	0.182	0.181	0.188	0.321	0.638	0.38	0.402	0.2	0.324	0.088																																																																																																																																																																																																																																																																																																	
CNBD	0.103	0.152	0.179	0.247	0.139	0.267	0.173	0.263	0.26	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																																	
ARAC1	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																																
ARAC2	0.104	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																															
MCDC	0.085	0.198	0.281	0.281	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																														
MCDD	0.085	0.198	0.281	0.281	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																													
RDLA	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																												
RDLK	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																											
RDLI	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																										
RDLH	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																									
RDLJ	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																								
RDL59	0.164	0.204	0.288	0.288	0.291	0.407	0.135	0.141	0.142	0.142	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	1.0	0.244	0.388	0.099	0.113	0.105	0.136	0.176	0.231	0.156	0.221	0.123	0.19	0.182	0.182	0.182	0.263	0.315	0.243	0.145	0.159	0.23																																																																																																																																																																																																																																																																																							
DCBA	0.109	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.089	0.08

Tabela 7.4: Matriz de similaridade usando a distância invertida, d^{-1} predita pelo modelo.

MOLECULE	DNP	LYD	CMG	CNBIPL	GLDC	OMCNB	PRT	TBC	THCP	THCVM2	CNBA	HHC	CNBD	ARAC1	ARAC2	MCDC	MCDI	RDLA	RDLC	RDLI	RDLI	RDLI	RDLS9	DCDA	DMDO	DMDE	DMCT	TCST	HD11	DDAA	MCRA	RBE
DNP	1.0	0.508	0.611	0.557	0.775	0.569	0.959	0.473	0.892	0.528	0.528	0.474	0.725	0.22	0.31	0.508	0.589	0.495	0.266	0.678	0.406	0.391	0.927	0.7	0.919	0.831	0.422	0.668	0.75	0.367	0.89	0.357
LYD	0.508	1.0	0.393	0.362	0.596	0.822	0.503	0.324	0.48	0.349	0.35	0.225	0.426	0.28	0.443	0.356	0.788	0.335	0.358	0.667	0.291	0.284	0.492	0.649	0.514	0.566	0.712	0.68	0.435	0.569	0.542	0.266
CMG	0.611	0.393	1.0	0.696	0.531	0.426	0.609	0.592	0.641	0.667	0.663	0.598	0.716	0.198	0.265	0.395	0.439	0.627	0.231	0.48	0.562	0.475	0.642	0.492	0.617	0.554	0.34	0.48	0.799	0.394	0.571	0.432
CNBIPL	0.557	0.362	0.696	1.0	0.48	0.392	0.563	0.757	0.596	0.908	0.91	0.76	0.706	0.188	0.249	0.362	0.491	0.816	0.219	0.441	0.599	0.566	0.578	0.45	0.547	0.501	0.316	0.436	0.684	0.284	0.522	0.499
GLDC	0.775	0.596	0.531	0.48	1.0	0.679	0.761	0.446	0.709	0.458	0.458	0.417	0.399	0.236	0.341	0.595	0.71	0.433	0.288	0.834	0.363	0.351	0.738	0.87	0.789	0.92	0.48	0.827	0.646	0.411	0.851	0.324
OMCNB	0.569	0.822	0.426	0.392	0.679	1.0	0.564	0.348	0.534	0.377	0.378	0.349	0.469	0.264	0.405	0.812	0.927	0.36	0.333	0.779	0.311	0.362	0.548	0.753	0.575	0.642	0.418	0.791	0.479	0.507	0.012	0.281
PRT	0.959	0.503	0.609	0.563	0.761	0.564	1.0	0.477	0.91	0.533	0.533	0.478	0.736	0.219	0.308	0.502	0.582	0.5	0.265	0.671	0.409	0.394	0.913	0.691	0.885	0.813	0.418	0.658	0.761	0.364	0.877	0.36
TBC	0.473	0.324	0.592	0.757	0.416	0.348	0.477	1.0	0.501	0.819	0.818	0.576	0.576	0.177	0.223	0.324	0.356	0.91	0.205	0.367	0.74	0.691	0.487	0.393	0.466	0.431	0.287	0.383	0.561	0.26	0.447	0.594
THCP	0.892	0.48	0.641	0.596	0.709	0.534	0.91	0.501	1.0	0.562	0.563	0.501	0.793	0.215	0.299	0.479	0.551	0.526	0.258	0.629	0.426	0.409	0.917	0.647	0.845	0.755	0.402	0.619	0.822	0.352	0.808	0.373
THCVM2	0.528	0.349	0.667	0.908	0.458	0.377	0.533	0.819	0.562	1.0	0.983	0.823	0.659	0.184	0.243	0.349	0.386	0.89	0.215	0.422	0.637	0.6	0.546	0.43	0.519	0.477	0.306	0.418	0.641	0.276	0.496	0.525
CNBA	0.528	0.35	0.663	0.91	0.458	0.378	0.533	0.818	0.563	0.983	1.0	0.821	0.66	0.184	0.243	0.349	0.386	0.887	0.215	0.423	0.636	0.6	0.546	0.431	0.519	0.477	0.306	0.418	0.641	0.276	0.496	0.525
HHC	0.474	0.235	0.598	0.76	0.417	0.349	0.478	0.976	0.501	0.823	0.821	1.0	0.577	0.177	0.231	0.426	0.482	0.609	0.242	0.541	0.479	0.458	0.758	0.554	0.705	0.632	0.364	0.533	0.956	0.322	0.667	0.413
CNBD	0.725	0.426	0.716	0.706	0.599	0.469	0.736	0.576	0.793	0.659	0.66	0.577	1.0	0.204	0.278	0.426	0.482	0.609	0.242	0.541	0.479	0.458	0.758	0.554	0.705	0.632	0.364	0.533	0.956	0.322	0.667	0.413
ARAC1	0.22	0.28	0.198	0.188	0.236	0.264	0.219	0.177	0.215	0.184	0.184	0.177	0.204	1.0	0.427	0.28	0.26	0.18	0.534	0.245	0.167	0.164	0.218	0.243	0.222	0.231	0.315	0.248	0.265	0.353	0.236	0.158
ARAC2	0.31	0.443	0.265	0.249	0.341	0.405	0.308	0.23	0.299	0.243	0.243	0.231	0.278	0.427	1.0	0.443	0.396	0.236	0.648	0.363	0.213	0.209	0.304	0.357	0.313	0.331	0.54	0.367	0.281	0.666	0.222	0.199
MCDC	0.568	0.556	0.395	0.362	0.595	0.812	0.502	0.324	0.479	0.349	0.349	0.225	0.426	0.28	0.443	1.0	0.786	0.335	0.358	0.664	0.291	0.283	0.492	0.647	0.515	0.566	0.711	0.68	0.434	0.569	0.541	0.265
MCDI	0.589	0.788	0.439	0.401	0.71	0.927	0.582	0.356	0.551	0.386	0.386	0.356	0.482	0.26	0.396	0.786	1.0	0.368	0.326	0.811	0.316	0.307	0.567	0.786	0.597	0.669	0.507	0.833	0.492	0.493	0.634	0.286
RDLA	0.495	0.335	0.627	0.516	0.433	0.36	0.5	0.91	0.526	0.89	0.887	0.916	0.609	0.18	0.209	0.401	0.692	0.647	0.512	0.409	0.488	0.45	0.295	0.397	0.593	0.267	0.67	0.467	0.562	0.467	0.562	
RDLC	0.266	0.558	0.231	0.219	0.288	0.333	0.265	0.295	0.258	0.215	0.215	0.205	0.242	0.534	0.648	0.358	0.326	0.209	1.0	0.304	0.191	0.188	0.261	0.3	0.268	0.281	0.418	0.306	0.244	0.491	0.275	0.18
RDLI	0.478	0.667	0.48	0.441	0.834	0.779	0.671	0.387	0.629	0.422	0.423	0.387	0.541	0.245	0.363	0.664	0.811	0.401	0.304	1.0	0.34	0.33	0.647	0.953	0.683	0.782	0.526	0.934	0.554	0.444	0.74	0.306
RDHI	0.406	0.291	0.502	0.599	0.363	0.311	0.409	0.74	0.426	0.637	0.636	0.738	0.479	0.167	0.213	0.291	0.316	0.692	0.191	0.34	1.0	0.892	0.417	0.346	0.401	0.475	0.261	0.338	0.469	0.239	0.387	0.749
RDLS	0.291	0.284	0.475	0.596	0.351	0.302	0.294	0.691	0.469	0.6	0.6	0.688	0.458	0.164	0.209	0.283	0.207	0.647	0.188	0.33	0.692	1.0	0.4	0.335	0.365	0.362	0.254	0.327	0.449	0.233	0.373	0.8
RDLS9	0.927	0.492	0.442	0.578	0.738	0.548	0.913	0.487	0.917	0.546	0.546	0.489	0.758	0.218	0.304	0.492	0.567	0.512	0.261	0.647	0.417	0.4	1.0	0.667	0.908	0.788	0.41	0.639	0.785	0.358	0.834	0.366
DCDA	0.7	0.649	0.492	0.45	0.87	0.753	0.691	0.393	0.647	0.43	0.431	0.394	0.554	0.243	0.357	0.647	0.786	0.409	0.3	0.953	0.346	0.335	0.667	1.0	0.706	0.812	0.514	0.924	0.568	0.435	0.765	0.31
DMDO	0.919	0.514	0.617	0.547	0.789	0.575	0.885	0.466	0.845	0.519	0.519	0.467	0.765	0.222	0.313	0.515	0.207	0.488	0.268	0.683	0.461	0.285	0.988	0.706	1.0	0.844	0.426	0.678	0.728	0.37	0.875	0.353
DMDE	0.831	0.566	0.554	0.501	0.92	0.642	0.813	0.431	0.755	0.477	0.477	0.432	0.632	0.231	0.331	0.566	0.609	0.45	0.281	0.782	0.375	0.362	0.788	0.812	0.844	1.0	0.461	0.772	0.651	0.396	0.914	0.333
DMCT	0.422	0.712	0.34	0.316	0.48	0.618	0.418	0.287	0.462	0.306	0.306	0.287	0.364	0.315	0.54	0.711	0.597	0.295	0.418	0.526	0.261	0.254	0.41	0.514	0.426	0.461	1.0	0.534	0.37	0.739	0.444	0.24
TCST	0.668	0.48	0.436	0.827	0.791	0.658	0.383	0.619	0.418	0.418	0.418	0.383	0.533	0.246	0.367	0.68	0.833	0.397	0.306	0.934	0.338	0.227	0.639	0.924	0.678	0.772	0.534	1.0	0.546	0.449	0.725	0.303
HD11	0.75	0.433	0.709	0.684	0.616	0.479	0.761	0.561	0.822	0.64	0.641	0.562	0.956	0.205	0.281	0.434	0.492	0.503	0.244	0.554	0.469	0.449	0.785	0.568	0.728	0.651	0.37	0.546	1.0	0.327	0.688	0.405
DDAA	0.367	0.569	0.394	0.284	0.411	0.507	0.364	0.26	0.352	0.276	0.276	0.261	0.322	0.353	0.666	0.569	0.493	0.267	0.491	0.444	0.239	0.233	0.358	0.435	0.37	0.396	0.739	0.449	0.327	1.0	0.384	0.221
MCRA	0.89	0.542	0.571	0.522	0.851	0.012	0.877	0.447	0.808	0.496	0.496	0.448	0.667	0.226	0.322	0.541	0.634	0.467	0.275	0.74	0.387	0.273	0.534	0.765	0.875	0.914	0.444	0.725	0.688	0.384	1.0	0.342
RBE	0.357	0.266	0.432	0.499	0.324	0.281	0.36	0.594	0.373	0.525	0.525	0.592	0.413	0.158	0.199	0.265	0.286	0.562	0.18	0.306	0.749	0.8	0.366	0.31	0.353	0.333	0.24	0.303	0.405	0.221	0.342	1.0

7.2 Visualizações dos resultados dos cálculos de estrutura eletrônica

Figura 7.3: Visualizações 3D dos orbitais de fronteira, MEPs e NCI das 32 moléculas.



7.3 Patente de software



INPI
INSTITUTO
NACIONAL
DA PROPRIEDADE
INDUSTRIAL
Assinado
Digitalmente

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DO DESENVOLVIMENTO, INDÚSTRIA, COMÉRCIO E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512024003481-1**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 25/01/2024, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: HeatLQC

Data de publicação: 25/01/2024

Data de criação: 25/01/2024

Titular(es): FUNDAÇÃO UNIVERSIDADE DE BRASILIA

Autor(es): JOÃO BATISTA LOPES MARTINS; ÉRICA CRISTINA MORENO NASCIMENTO; RAFAEL DO COUTO CAMPOS DE JESUS

Linguagem: PYTHON

Campo de aplicação: BL-08

Tipo de programa: TC-01

Algoritmo hash: SHA-512

Resumo digital hash:

9FE6F754E1C619A0899ED95433CFCBCF4CF3F1EF64CF1A0CD9AC73E2F1114AAEC6BF681DCAD5F012194556
0EE7C0916F945194597B72D024683C0C868658294

Expedido em: 01/10/2024

Aprovado por:

Carlos Alexandre Fernandes Silva
Chefe da DIPTO