

1 2 9 0



UNIVERSIDADE D
COIMBRA

Soraia Filipa Drysdale Assis

**CHARACTERIZING THE EMERGING
PSYCHOACTIVE DRUGS BY MINING DATA
FROM CHEMICAL DATABASES AND
MOLECULAR SIMULATIONS**

Dissertação no âmbito do Projeto Científico do Mestrado em
Química Forense orientada pela Professora Doutora Sandra
Cristina da Cruz Nunes e Professora Doutora Tânia Firmino
Guerra Guerreiro Cova e apresentada ao Departamento de
Química da Faculdade de Ciências e Tecnologia da Universidade
de Coimbra.

Setembro de 2021

Soraia Filipa Drysdale Assis

CHARACTERIZING THE EMERGING PSYCHOACTIVE
DRUGS BY MINING DATA FROM CHEMICAL
DATABASES AND MOLECULAR SIMULATIONS

Dissertação no âmbito do Projeto Científico do Mestrado em Química Forense orientada
pela Professora Doutora Sandra Cristina da Cruz Nunes e Professora Doutora Tânia Firmino
Guerra Guerreiro Cova e apresentada ao Departamento de Química da Faculdade de
Ciências e Tecnologia da Universidade de Coimbra.

Setembro de 2021



UNIVERSIDADE D
COIMBRA

**“Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less.”**

- Marie Curie

Agradecimentos

No meu primeiro dia em Coimbra disseram-me “quem chega a Coimbra, já não quer mais sair”. Todos os dias do meu percurso académico fizeram com que essa frase fosse a maior verdade de todas. Mas tudo tem um fim. Não é um fim que me deixe particularmente triste, pelo contrário! A memória de todas as emoções que tive a oportunidade de viver nesta bela cidade, fez com que todos os esforços valessem a pena. Desde pequena que olhava para a televisão e dizia “quero ser a Sílvia do Inspetor Max”. Agora posso dizer que tenho o requisito mínimo, por isso veremos o que o futuro me reserva. Ao longo deste percurso tive o privilégio de contar com pessoas extraordinárias e não podia de terminar este percurso sem prestar aqui os devidos agradecimentos.

Às minhas orientadoras, Doutora Sandra Nunes e Doutora Tânia Firmino, um grande obrigada pela paciência, orientação e apoio. Mais ainda, pela confiança que demonstraram em que era capaz de ultrapassar os desafios que me propunham.

A quem de bom Coimbra me trouxe, tenho de agradecer pelo apoio incondicional, convívio e sinceridade. À Andreia Melo e Patrícia por terem aturado as minhas ideias malucas sem dizerem não. À Bruna e Joana que aturaram as minhas crises existenciais em casa. À Andreia Amaro por ter não me deixar baixar os braços e pela lealdade.

A quem já fazia parte da minha vida, obrigado por ainda hoje permanecerem nela, mesmo que por vezes eu desaparecesse. À Beatriz, Catarina e Filipa que sempre me apoiaram e mostraram o orgulho que têm em mim. À Ana Catarina que mesmo passados 22 anos ainda caminha ao meu lado. À Catarina e Rita pela lealdade e me apoiarem incondicionalmente. E à Beatriz e Catarina Silva por não me deixarem baixar os braços e por insistirem que as metas têm de ser alcançadas.

À minha mãe, que apesar dos tempos difíceis que viveu e continua a viver não larga o sorriso que tem na cara e, muito menos, o otimismo e desenrasque que me surpreende a cada dia.

À minha avó, minha melhor amiga, que com os tempos difíceis que estamos a viver não tenho a sorte de ter o seu abraço. Mas não é por isso que deixa de transmitir a sua preocupação e amor que tem por mim.

Ao meu pai, a razão de eu ter chegado ao fim desta viagem. Não tenho palavras para agradecer tudo o que fez, e ainda faz, por mim e não terei dias de vida nem atos suficientes para agradecer por tudo.

Obrigada, também, a todos aquele que de uma maneira ou de outra tive o privilégio de me cruzar nesta viagem. Mas acima de tudo, Obrigada Coimbra por me receberes como se te pertencesse.

“Coimbra da saudade entre o sonho e a realidade”

Resumo

O aparecimento e consumo de Novas Substâncias Psicoativas, NSP, tem vindo a crescer consideravelmente a nível mundial, constituindo por isso uma preocupação para os vários organismos internacionais responsáveis pelo acompanhamento e regulação da problemática das drogas de consumo. Na categoria das NSP incluem-se todos os compostos, recém-sintetizados, ou recentemente disponíveis no mercado e ainda os usados de forma impropria, com propriedades psicotrópicas, puras ou misturas, que não sejam controlados pelas Convenções das Nações Unidas sobre os estupefacientes e substâncias psicotrópicas. Estas substâncias são criadas para mimetizar os efeitos das drogas já existentes e controladas no âmbito das referidas convenções e que possam constituir ameaça para a saúde pública.

Atendendo ao aumento expressivo da comercialização e consumo das NSP a investigação permanente deste fenómeno bem como a rápida sinalização e caracterização deste tipo de compostos constitui uma necessidade cada vez mais premente. Neste sentido, o trabalho aqui apresentado visa contribuir para a criação de novas ferramentas e metodologias que permitam caracterizar de forma rápida e eficiente NSP desconhecidas tendo por base o seu perfil químico e a comparação com NSP já regulamentadas.

A combinação de dois métodos não supervisionados de aprendizagem computacional, a Análise de Componentes Principais (PCA) e a Análise Hierárquica de agrupamentos (HCA), permitiu determinar o conjunto de descritores moleculares mais relevantes para a discriminação de famílias de NSP e ainda a formação de grupos com características estruturais comuns.

Os padrões de interação mais relevantes característicos destes grupos foram explorados recorrendo à simulação por Dinâmica Molecular e *Docking*. Selecionando como grupo modelo a família dos Canabinoides Sintéticos foi possível verificar que compostos pertencentes ao mesmo grupo apresentam perfis estruturais e comportamento semelhante quer na presença de uma membrana modelo, quer na interação com o recetor canabinoide tipo 1.

Os resultados obtidos permitem estabelecer uma metodologia de classificação eficaz para NSP não regulamentadas sem necessidade de recorrer a estudos experimentais.

Palavras-chave: NSP, Aprendizagem Computacional, Dinâmica Molecular, *Docking*.

Abstract

The emergence and consumption of New Psychoactive Substances, NPS, has been growing considerably worldwide, and is therefore a concern for the various international entities responsible for monitoring and regulating the problem of drugs abuse. The classification as NPS includes all compounds, newly synthesized or recently available on the market, as well as those used improperly, with psychotropic properties, either pure or in mixtures, which are not controlled by the United Nations Conventions on Narcotic Drugs and Psychotropic Substances. These substances are created to mimic the effects of drugs already existing and controlled under these conventions, and may pose a threat to public health.

Due to the significant increase in the commercialization and consumption of NPS, the investigation of this phenomenon, as well as the rapid identification and characterization of this type of compounds, is an increasingly pressing need.

In this sense, the work presented here aims to contribute to the development of new tools and methodologies to efficiently characterize unknown NPS based on their chemical profile and comparison with already regulated NPS.

The combination of two unsupervised machine learning methods, Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA), allowed the determination of the most relevant molecular descriptors for the discrimination between families of NSP and also the formation of groups with common structural characteristics.

The most relevant interaction patterns of these groups were explored using Molecular Dynamics and Docking simulation. Selecting as model group the family of synthetic cannabinoids, it was possible to verify that compounds belonging to the same group present similar structural profiles and behavior both in the presence of a model membrane and in the interaction with the cannabinoid receptor type 1.

The results obtained allow establishing an effective classification methodology for unregulated NPS without the need for experimental studies.

Keywords: NPS, Machine Learning, Molecular Dynamics, Docking

Índice

Agradecimentos	iii
Resumo.....	v
Abstract	vii
Índice	ix
Abreviaturas	xi
Índice de Figuras.....	xiii
Índice de Tabelas.....	xix
Capítulo 1 Introdução.....	1
1.1 Substâncias Psicoativas	5
1.2 Novas Substâncias Psicoativas	5
1.2.1 Aminoindanos.....	9
1.2.2 Canabinoides Sintéticos	10
1.2.3 Catinonas Sintéticas	11
1.2.4 Fenetilaminas	12
1.2.5 Novas Benzodiazepinas	13
1.2.6 Opióides Sintéticos	14
1.2.7 Piperazinas	14
1.2.8 Substâncias tipo Fenciclidinas	15
1.2.9 Triptaminas.....	15
1.3 Objetivos	16
1.4 Aplicação de métodos computacionais na caracterização de drogas.....	16
Capítulo 2 Teoria e Métodos	19
2.1 Quimiometria e aprendizagem computacional aplicado à Química Forense	21
2.2 Análise Hierárquica de Agrupamentos.....	23
2.3 Análise de Componentes Principais	25
2.3.1 Passos na aplicação do PCA.....	26
2.4 Modelação e Simulação Molecular	32
2.4.1 Dinâmica Molecular.....	33
2.4.2 <i>Docking</i> molecular	39
2.4.3 Método do Gradiente Independente	41
Capítulo 3 Descrição da base de dados.....	43
3.1 Coordenadas 3D das NPS	45
3.2 Descritores Moleculares.....	71
Capítulo 4 Discriminação entre NSP com base na estrutura química.....	73

4.1 Seleção dos descritores moleculares	75
4.2 Seleção das NSP baseada na similaridade estrutural.....	83
Capítulo 5 Detalhes das simulações moleculares	95
5.1 Dinâmica Molecular.....	97
5.2 <i>Docking Molecular</i>	98
Capítulo 6 Caracterização dos padrões de interação.....	103
6.1 Interação entre as NSP e uma membrana modelo	105
6.2 Interação entre as moléculas de NSP e o recetor CB ₁	110
Capítulo 7 Conclusão	121
Referências bibliográficas.....	125
Anexos	141
Anexo 1.....	143
Anexo 2.....	146
Anexo 3.....	148
Anexo 4.....	151

Abreviaturas

ADT	AutoDockTols
BBB	Blood-Brain-Barrier - Barreira hematoencefálica
CB₁	Recetor do Canabinoide tipo 1
CB₂	Recetor do Canabinoide tipo 2
DM	Dinâmica Molecular
EMCDDA	European Monitoring Centre for Drugs and Drug Addiction
HCA	Hierarchical Clustering Analysis – Análise Hierárquica de Agrupamentos
IGM	Independent Gradient Model
k	Grupos
MSD	Mean Square Displacement – Perfil de deslocamento médio
NSP	Novas Substâncias Psicoativas
PC	Principal Components – Componentes Principais
PCA	Principal Component Analysis – Análise de Componentes Principais
SNC	Sistema Nervoso Central
S_{co}	Parâmetro de ordem
Δ⁹-THC	(-) - trans-Δ ⁹ -tetrahydrocannabinol

Índice de Figuras

Figura 1.1 Representação gráfica do número de NSP notificadas pela primeira vez em cada ano. Estes dados estatísticos foram retirados dos relatórios anuais sobre drogas entre 2006 e 2020. ¹⁹⁻³³	7
Figura 1.2 Alguns marcos importantes na história das NSP entre 2005 e 2020. As informações destacadas neste esquema foram retirados dos Relatórios Anuais sobre droga emitidos pelo EMCDDA entre 2006 e 2020. ¹⁹⁻³³	8
Figura 1.3 Estrutura da anfetamina a) e da 2-aminoindane b). ³⁶	9
Figura 1.4 Modelo estrutural proposto pelo EMCDDA e os respetivas possibilidades de grupos substituintes para obtenção de diferentes Canabinoides Sintéticos (adaptado de ⁴¹).	11
Figura 1.5 Base da estrutura molecular das Catinonas Sintéticas, sendo R ¹ , R ² , R ³ , R ⁴ e R ⁵ as possíveis posições para grupos substituintes. ⁴⁵	12
Figura 1.6 Estrutura molecular das Fenetilaminas, sendo que de R ¹ a R ⁸ correspondem às possíveis posições para grupos substituintes. ⁴⁷	12
Figura 1.7 Estruturas moleculares base das Novas Benzodiazepinas, sendo que de R ¹ a R ¹³ correspondem às possíveis posições para grupos substituintes. ⁵⁰	13
Figura 1.8 Exemplos de dois compostos pertencentes à família dos Opióides Sintéticos: Fentanil (A) e U-47700 (B). ⁵²	14
Figura 1.9 Estrutura molecular base das duas classes da família das Piperazinas, derivados de Benzilpiperazina (A) e derivados de Fenilpiperazina (B). ⁵⁴	14
Figura 1.10 Estrutura molecular do PCP.	15
Figura 1.11 Estrutura base das triptaminas. ⁵⁷	15
Figura 1.12 Classificação das Triptaminas (adaptado de ⁵⁶).	15
Figura 2.1 Exemplo de uma representação gráfica obtida pelo HCA, o dendrograma.	23
Figura 2.2 Representação esquemática do processo de redução da dimensionalidade.....	29
Figura 2.3 Representação gráfica dos valores próprios em função do número de componentes consideradas (adaptado de ⁷⁸).	31
Figura 2.4 Esquema representativo das etapas principais da Dinâmica Molecular (DM) (adaptado de ⁸⁷).	34
Figura 2.5 Diagrama representativo dos quatro tópicos necessários para desenvolver um modelo de simulação molecular (adaptado de ⁸⁹).....	36

Figura 2.6 Esquema representativo do efeito da imposição de Condições de Fronteira Periódicas em duas dimensões (adaptado de ⁸⁸).....	37
Figura 2.7 Esquema-resumo dos pontos fundamentais no estudo do <i>Docking</i> molecular (adaptado de ⁹⁴).	40
Figura 2.8 Diagrama representativo do conjunto de etapas no <i>Docking</i> molecular (adaptado de ⁹³).	41
Figura 4.1 <i>Screeplot</i> para o conjunto de dados inicial constituído por 301 moléculas e 66 descritores moleculares.	77
Figura 4.2 <i>Screeplot</i> para o conjunto de dados formado por 301 moléculas e 27 descritores moleculares.	79
Figura 4.3 <i>Screeplot</i> para o conjunto de dados formado por 301 moléculas e 43 descritores.	81
Figura 4.4 Representação gráfica da posição das variáveis relativamente às duas primeiras componentes principais.	81
Figura 4.5 Relação entre os descritores e as duas primeiras componentes principais.	82
Figura 4.6 Representação na forma de biplot resultante da aplicação da análise de componentes principais sobre o conjunto de dados constituído pelas 301 moléculas e 43 descritores moleculares em função das duas primeiras componentes principais, PC1 e PC2, com uma recuperação de 41,2%.	82
Figura 4.7 Dendrograma resultante da análise HCA usando a distância Euclidiana e método <i>Ward</i> sobre o conjunto de dados formado pelas 301 moléculas e 43 descritores moleculares. As cores das abreviaturas correspondem à família na qual as NSP estão inseridas.	84
Figura 4.8 Representação gráfica do método <i>Silhouette</i> para o conjunto de dados de 234 moléculas e 43 descritores moleculares.	86
Figura 4.9 Dendrograma resultante da análise HCA usando a distância Euclidiana e método <i>Ward</i> sobre o conjunto de dados formado pelas 234 moléculas e os 43 descritores. As cores das abreviaturas correspondem à família na qual as NSP estão inseridas. As cores das ramificações corresponde ao grupo resultante da análise do HCA.	86
Figura 4.10 Visão detalhada do primeiro grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 48 NSP da família das Piperazinas, Catinonas Sintéticas, Fenetilaminas, Triptaminas, Aminoindanes e Outras drogas.	87

Figura 4.11 Exemplos de pares de NSP Controlo – NSP Não ilícita das famílias das Piperazinas, Catinonas Sintéticas e Fenetilaminas presentes no primeiro grupo do dendrograma da Figura 4.9.....	87
Figura 4.12 Visão detalhada do segundo grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 49 NSP da família das Triptaminas, Fenetilaminas, Catinonas Sintéticas, Aminoindanes e Outras drogas.....	88
Figura 4.13 Exemplos de pares de NSP Controlo – NSP Não ilícita das famílias das Triptaminas e Fenetilaminas presentes no segundo grupo do dendrograma da Figura 4.9.	88
Figura 4.14 Visão detalhada do terceiro grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 70 NSP das famílias dos Opióides Sintéticos, Piperazinas, Substâncias tipo Fenciclidinas, Fenetilaminas, Catinonas Sintéticas, Triptaminas, Novas Benzodiazepinas, Canabinoides Sintéticos e Outras drogas.....	89
Figura 4.15 Exemplos de pares de NSP Controlo – NSP Não ilícita das famílias das Substâncias tipo Fenciclidinas, Opióides Sintéticos, Catinonas Sintéticas e Fenetilaminas presentes no terceiro grupo do dendrograma da Figura 4.9.....	89
Figura 4.16 Visão detalhada do quarto grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 26 NSP das famílias dos Novas Benzodiazepinas e Outras drogas.	90
Figura 4.17 Exemplos de um par Controlo – NSP Não ilícita da família das Novas Benzodiazepinas presente no quarto grupo do dendrograma da Figura 4.9.....	90
Figura 4.18 Aproximação ao quinto grupo do dendrograma presente na Figura 4.9. Neste grupo estão presentes 18 NSP da famílias dos Canabinoides Sintéticos.....	90
Figura 4.19 Exemplo de um par de NSP Controlo – NSP Não ilícita da família dos Canabinoides Sintéticos presente no quinto grupo do dendrograma da Figura 4.9.....	91
Figura 4.20 Visão detalhada do sexto grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 24 NSP das famílias dos Canabinoides Sintéticos, Opióides Sintéticos, Triptaminas, Novas Benzodiazepinas e Outras drogas.	91
Figura 4.21 Exemplo de um par NSP Controlo – NSP Não ilícita da família dos Canabinoides Sintéticos presente no sexto grupo do dendrograma da Figura 4.9.	92
Figura 4.22 Representação na forma de biplot resultante da aplicação da análise de componentes principais sobre o conjunto de dados constituído pelas 234 moléculas e 43 descritores moleculares em função das duas primeiras componentes principais, PC1 e PC2, com 27,8% e 12,3% de variabilidade recuperada.	93

Figura 5.1 Estrutura molecular das NSP selecionadas para o estudo a nível molecular por simulação computacional. As NSP selecionadas pertencem à família dos Canabinoides Sintéticos.	97
Figura 6.1 Estrutura química do lípido 1-palmitoil-2-oleoil-sn-glicero-3-fosfocolina (POPC) constituinte da membrana modelo utilizada no estudo por DM.	105
Figura 6.2 Perfis de densidade de (A) CS12_C (JWH-018), (B) CS33_C (5F-AMB-PINACA), (C) CS55_NI (5F-EBM-PINACA) através da bicamada de POPC (1 NSP:128 POPC). As moléculas de POPC foram identificadas pelos átomos P (vermelho) e N (amarelo) dos grupos polares e pelos átomos dos grupos metilo terminais (verde). O solvente (água) encontra-se representado pelo perfil azul-claro. As NSP foram representadas pelos respetivos centros de massa (azul, roxo e cor-de-rosa, respetivamente). De notar que a coordenada Z representa o corte longitudinal da membrana.	106
Figura 6.3 Configurações representativas do posicionamento das moléculas de NSP estudadas na bicamada de POPC. (A) CS12_C. (B) CS33_C. (C) CS55_NI	108
Figura 6.4 Resultados da análise de DM: (A) perfil de deslocamento quadrático médio (MSD) da membrana POPC na ausência e na presença das moléculas de NSP; (B) Parâmetro de ordem de deutério para a cauda saturada (S_{cd}), estimada ao longo da bicamada na ausência e presença das NSP; (C) Parâmetro de ordem de deutério para a cauda insaturada (S_{cd}), estimada ao longo da bicamada na ausência e na presença das NSP.	109
Figura 6.5 Conformação representativa das interações mais prevalentes estabelecidas entre o recetor CB ₁ e o canabinoide CS12_C resultante da simulação por <i>docking</i> molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) <i>Poseview</i> do ligando CS12_C mostrando a ligação de hidrogénio estabelecida com o resíduo Ser383 e a ligação π - π com o resíduo Phe170, obtida com recurso ao servidor <i>Protein Plus</i> (https://proteins.plus/); (B) Diagrama bidimensional no qual estão sumariadas os tipos de interação mais prováveis entre o CS12_C e os resíduos do recetor CB ₁ . Resíduos rosa – interações hidrofóbicas, laranja - interações eletrostáticas, verde-claro - van der Waals, verde - ligações de hidrogénio convencionais; (C) Conformação mais estável do complexo (<i>Dock pose</i>) entre CS12_C (estrutura a preto) e do recetor (roxo) correspondente à classificação (<i>score</i>) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao <i>software Discovery Studio Visualizer</i>	114

Figura 6.6 Conformação representativa de uma das interações mais favorável entre o recetor CB1 e o canabinoide CS33_C resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS33_C mostrando a ligação de hidrogénio estabelecida com o resíduo His178 e a ligação π - π com o resíduo Phe170, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional no qual estão sumariadas os tipos de interação mais prováveis entre o CS33_C e os resíduos do recetor CB₁. Resíduos rosa – interações hidrofóbicas, verde-claro - van der Waals, verde - ligações de hidrogénio convencionais; (C) Conformação do complexo (*Dock pose*) entre CS33_C (estrutura a preto) e do recetor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*. 115

Figura 6.7 Conformação representativa de uma das interações mais favorável entre o recetor CB1 e o canabinoide CS33_C resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS33_C mostrando a ligação de hidrogénio estabelecida com o resíduo Ser383 e a ligação π - π com o resíduo Phe174, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional sumariando os tipos de interação mais prováveis entre o CS33_C e os resíduos do recetor CB₁. Resíduos rosa – interações hidrofóbicas, verde-claro – ligações de hidrogénio não convencionais, verde - ligações de hidrogénio convencionais, roxo – interações π - σ , azul – ligação com o halogénio; (C) Conformação do complexo (*Dock pose*) entre CS33_C (estrutura a preto) e do recetor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*. 116

Figura 6.8 Conformação representativa da interações mais favorável entre o recetor CB1 e o canabinoide CS55_NI resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS55_NI mostrando as ligações de hidrogénio estabelecidas com os resíduos His178 e Ser383 e a ligação π - π com o resíduo Phe170, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional no qual estão sumariadas os tipos de interação mais prováveis entre o CS55_NI e os resíduos do recetor CB₁.

Resíduos rosa – interações hidrofóbicas, verde-claro – van der Waals, verde - ligações de hidrogénio convencionais, azul – ligação com o halogénio; (C) Conformação do complexo (*Dock pose*) entre CS55_NI (estrutura a preto) e do recetor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*..... 117

Figura 6.9 Visão geral dos padrões de interação dos complexos formados entre os canabinoides e o recetor CB1, considerando a natureza e a força das interações intermoleculares identificadas nas conformações de menor energia obtidas nas simulações por docking molecular. (A) CS12_C:CB1, (B) CS33_C:CB1 e (C) CS55_NI:CB1. As intensidades dos picos observadas nos gráficos 2D estão relacionadas com as energias de estabilização do complexo . Os contactos mais prevalentes entre os canabinoides e o recetor são também ilustrados pelas zonas de isosuperfície em 3D (para os compostos CS33_C e CS55_NI). Os sistemas presentes nas imagens dos esquemas (B) e (C) mostram as ligações de hidrogénio entre os compostos CS33_C e CS55_NI com os resíduos His178 e Ser383, respetivamente. O esqueleto molecular dos compostos CS33_C e CS55_NI são representados de acordo com os seus átomos carbono, hidrogénio, oxigénio, azoto e fluor, pelas cores preto, branco, vermelho, azul e rosa, respetivamente. Os resíduos His178 e Ser383 estão representados pelas cores azul e laranja, repetivamente..... 118

Índice de Tabelas

Tabela 2.1 Sumário dos critérios de ligação mais conhecidos e o respetivo dendrograma....	24
Tabela 2.2 Diferentes tipos de movimento presente em diversos sistemas e respetivos tempos de integração (adaptado de ⁸⁸).	35
Tabela 3.1 Lista de drogas pertencentes à família dos Aminoindanes divididos pelas categorias Controlo (C) e Não ilícitas (NI).	46
Tabela 3.2 Lista de drogas pertencentes à família dos Canabinoides Sintéticos divididos pelas categorias Controlo (C) e Não ilícitas (NI).	46
Tabela 3.3 Lista de drogas pertencentes à família das Catinonas Sintéticas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	52
Tabela 3.4 Lista de drogas pertencentes à família das Fenetilaminas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	55
Tabela 3.5 Lista de drogas pertencentes à família das Novas Benzodiazepinas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	58
Tabela 3.6 Lista de drogas pertencentes à família dos Opióides Sintéticos divididos pelas categorias Controlo (C) e Não ilícitas (NI).	62
Tabela 3.7 Lista de drogas pertencentes à família das Piperazinas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	64
Tabela 3.8 Lista de drogas pertencentes à família das Substâncias tipo Fenciclidinas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	65
Tabela 3.9 Lista de drogas pertencentes à família das Triptaminas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	66
Tabela 3.10 Lista de drogas pertencentes à família das Outras drogas divididas pelas categorias Controlo (C) e Não ilícitas (NI).	69
Tabela 3.11 Exemplos de alguns descritores moleculares e a respetiva categoria.	72
Tabela 4.1 Lista dos 66 descritores moleculares selecionados após o pré-processamento. ..	75
Tabela 4.2 Valores próprios e percentagem de recuperação da informação inicial em relação ao número de componentes principais para o conjunto de dados inicial constituído por 301 moléculas e 66 descritores moleculares.	76
Tabela 4.3 Lista dos 27 descritores moleculares selecionados após a aplicação do critério definido pela expressão 4.1.	78

Tabela 4.4 Valores próprios e percentagem de recuperação da informação para o conjunto de dados constituído por 301 moléculas e 27 descritores moleculares.....	78
Tabela 4.5 Lista dos 53 descritores moleculares após serem adicionados os descritores de ADMET.....	79
Tabela 4.6 Lista dos 43 descritores moleculares selecionados após a aplicação do critério definido pela expressão 4.1.	80
Tabela 4.7 Valores próprios e percentagem de informação recuperada para o conjunto de dados contendo 301 moléculas e 43 descritores moleculares.....	80
Tabela 4.8 Lista das 67 NSP eliminadas do conjunto de dados inicial.	85
Tabela 5.1 Principais publicações identificando os resíduos do recetor CB1 responsáveis pela interação com moléculas da família dos Canabinoides Sintéticos.	99
Tabela 6.1 Valores de energia de interação, E_{int} (Kcal/mol), correspondentes às configurações dos complexos entre as NSP selecionadas e o recetor CB ₁ obtidos em cada simulação. dp corresponde ao desvio padrão.....	110
Tabela 6.2 Lista de resíduos e interações responsáveis pela estabilização dos complexos entre o recetor CB ₁ e as NSP, identificados por <i>docking</i> molecular. As interações identificadas incluem interações hidrofóbicas, ligações de hidrogénio convencionais e não convencionais. Os resíduos marcados a negrito correspondem aos identificados nas três simulações realizadas para cada molécula. Os resíduos assinalados a cor rosa correspondem a resíduos que são comuns aos Canabinoides CS33_C e CS55_NI e que não surgem no Canabinoide CS12_C para uma determinada interação. Por outro lado, os resíduos de cor azul referem-se a resíduos que interagem apenas com o compostos CS12_C.	111
Tabela A 1 Descritores presentes na categoria dos Descritores Híbridos. Nesta categoria existem 2 classes de descritores.	143
Tabela A 2 Descritores presentes na categoria dos Descritores Constitucionais. Nesta categoria existem 16 classes de descritores.	143
Tabela A 3 Descritores presentes na categoria dos Descritores Eletrónicos. Nesta categoria existem 5 classes de descritores.	143
Tabela A 4 Descritores presentes na categoria dos Descritores Topológicos. Nesta categoria existem 25 classes de descritores.	144
Tabela A 5 Descritores presentes na categoria dos Descritores Geométricos. Nesta categoria existem 3 classes de descritores.	145

Tabela A 6 Definição dos descritores pertencentes ao conjunto final.	146
Tabela A 7 Lista de resíduos do recetor que interagiu com os compostos em estudo em cada artigo. Os resíduos selecionados para o estudo estão representados pela cor rosa e serão utilizados para definir a caixa de simulação no método <i>Docking</i>	148
Tabela A 8 Valores das coordenadas cartesianas usadas para a determinação do centro de caixa de simulação pelo método <i>Docking</i>	151

Capítulo 1

Introdução

1. Introdução

A Toxicologia é o ramo da ciência que procura estudar o comportamento e efeitos adversos de substâncias dentro do organismo. Nesta simples definição há dois termos bastante importantes quando nos referimos a esta área, as substâncias e o organismo.¹ As substâncias, que também podem ser denominadas por xenobióticos, referem-se a qualquer composto que seja estranho ao organismo.² Já o organismo refere-se a qualquer ser vivo, como por exemplo o ser humano, e que irá ser o alvo destas substâncias.¹ A Toxicologia Forense, mais especificamente, é o ramo que aplica os conhecimentos da toxicologia entre outros na análise toxicológica *post-mortem* para determinação da causa de morte, e na monitorização de drogas numa perspetiva de avaliação da dinâmica de compostos usados ilicitamente e/ou regulamentados.^{3,4} Um dos cientistas mais notórios nesta área foi Paracelsus e é conhecido pela sua célebre frase *“All substances are poisons; there is none which is not a poison. The right dose differentiates a poison from a remedy”*, com isto, Paracelsus afirmava que qualquer composto possui uma determinada dose que pode ser letal para o organismo, inclusive os compostos que são utilizados como medicamento.⁵ No entanto, este não foi o único toxicologista forense a receber méritos na história nesta área, o primeiro cientista a ser reconhecido foi Mathieu Joseph Bonaventure Orfila que em 1813 lançou um livro cujo título era *“Traité des poisons tirés des règnes minéral végétal et animal”*. Este cientista é considerado como um dos pioneiros na inserção do conhecimento da química em sede judicial.²

O uso das ferramentas da Química Analítica na Toxicologia Forense permite estimar qualitativa e/ou quantitativamente as substâncias em estudo. Para tal são utilizados métodos de triagem, preparação e análise da amostra. Para estas análises existem valores de referência já definidos, ou seja, valores estabelecidos para determinadas substâncias, classe de substâncias e metabolitos. Estes valores podem variar com a matriz utilizada, limite de deteção da técnica, entre outros aspetos. Os métodos analíticos mais comuns nesta área são os Testes Colorimétricos, Cromatografia em Camada Fina (TLC), Espectroscopia de Absorção no Ultravioleta e Visível (UV/VIS), Espectroscopia de infravermelho (IV), Cromatografia em Fase Gasosa acoplada à Espectrometria de Massa (GC-MS), Cromatografia em fase líquida (LC), entre outros.⁶

Na verdade, o consumo de compostos que hoje sabemos serem tóxicos e ilícitos não é recente. O seu uso teve início nos primórdios da evolução humana, quando o conhecimento dos efeitos tóxicos destas substâncias era inexistente. O que comprova esta afirmação, é o facto de terem sido encontradas substâncias psicoativas em grutas que tinham sido ocupadas por tribos pré-históricas. Estas substâncias foram encontradas junto a objetos que estariam ligados a cerimónias sagradas⁷ e eram utilizadas como elo de ligação entre a realidade e a “vida prometida”.⁸ Os fundadores da medicina científica, Hipócrates e Galene, definiam a droga como todo e qualquer composto que teria a capacidade de produzir efeitos nocivos ao corpo humano. Os compostos que, aparentemente, não produziam efeitos nocivos eram utilizadas como fonte de prazer e faziam parte de tratamento médico. Por exemplo, na Antiga Grécia, o ópio era utilizado na área da medicina como remédio.⁸ Por volta do século XVIII deu-se início à comercialização de substâncias psicoativas. O número de consumidores de ópio cresceu de tal modo, que o governo de Pequim proibiu o seu comércio, dando origem às guerras do ópio.⁷ No ano de 1860 foi sintetizada, pela primeira vez a cocaína. Este composto foi estudado por Freud que, no ano de 1884 lançou, o seu ensaio “Über Coca”. Após esta publicação houve um aumento de prescrições médicas da cocaína para o tratamento da ansiedade e depressão. No final deste século, o consumo da cocaína, do ópio e dos respetivos alcaloides aumentou especialmente em famílias com elevadas condições económicas e mais cultas.⁸ Com o aumento do comércio de estupefacientes, houve a necessidade de proibir este comércio, através de uma série de conferências internacionais, como por exemplo a de Xangai em 1909 e a de Genebra em 1931.⁷ As anfetaminas chegaram ao mercado das drogas no século XX, na década de trinta, especialmente no período da II Guerra Mundial. Contudo, foi com o movimento hippie, na década de sessenta, que estas drogas começaram a ser consumidas pela população em geral, não olhando a classes sociais e aumentando o número de jovens consumidores. Duas décadas depois chegaram a este mercado as drogas sintéticas, que eram sintetizadas em laboratórios ilegais a uma velocidade elevada. Além de este grupo de drogas ser muito vasto, as suas propriedades são muitas das vezes desconhecidas e os seus efeitos imprevisíveis.⁸

O foco deste estudo é exatamente este último grupo de drogas sintéticas, particularmente as que não estão regulamentadas.

1.1 Substâncias Psicoativas

As substâncias psicoativas atuam no cérebro, afetando a sua atividade através de alterações na percepção, estado emocional, grau de consciência, atividade cognitiva e comportamento.⁹ O principal alvo destas substâncias é o sistema nervoso central (SNC) e é ao entrarem em contacto com o seu alvo que produzem os seus efeitos. Existem diferentes maneiras de uma substância psicoativa ser consumida, as substâncias podem entrar no organismo através da via oral, inalatória e injetada no sangue, músculo ou cérebro. O modo mais seguro de as consumir é através da via oral, contudo se forem injetadas diretamente no cérebro o efeito é muito mais rápido.¹⁰ A comunicação ao cérebro é feita através dos neurónios e o contacto entre substâncias psicoativas com os neurónios pode prejudicar esta comunicação. Os efeitos provocados pelo consumo destas substâncias envolvem a libertação de neurotransmissores ou de sistemas de segundos mensageiros. Além disso, a presença destes compostos no organismo podem promover alterações nas células.¹¹ A metabolização é feita nos rins, fígado e na bile. Estes compostos são eliminados através da urina, fezes, suor e ar expirado.¹⁰ Curiosamente, uma das substâncias psicoativas mais consumida nos dias de hoje é a cafeína.¹¹

O aumento drástico do consumo destas substâncias levou os países pertencentes às Nações Unidas a definir um conjunto de medidas de modo a diminuir o tráfico ilícito das substâncias incluídas nas listas dos documentos redigidos pela Convenção Única das Nações Unidas sobre Estupefacientes de 1961, que foi posteriormente alterada pelo Protocolo de 1972, ou pela Convenção das Nações Unidas sobre as Substâncias Psicotrópicas de 1971.^{12,13}

1.2 Novas Substâncias Psicoativas

As Novas Substâncias Psicoativas (NSP) são definidas como todos os compostos que na sua forma pura ou em preparação, não são controladas pela Convenção Única das Nações Unidas sobre Estupefacientes de 1961 ou pela Convenção das Nações Unidas sobre as Substâncias Psicotrópicas de 1971, mas que constituem uma ameaça para a saúde pública quando comparadas com as substâncias que são controladas.¹⁴ No entanto, outras organizações utilizam definições ligeiramente diferentes, como é o caso da *International Narcotics Control Board* que considera que NSP as substâncias que foram desenvolvidas com o propósito de não serem detetadas nas medidas de controlo existentes para as drogas já

regulamentadas. Assim, estas substâncias são sintetizadas com base na estrutura química das drogas clássicas, contendo pequenas modificações, dando origem a um novo composto com efeitos farmacológicos semelhantes ao das substâncias já controladas.¹⁵

Este tipo de substâncias são muitas vezes designados por “*designer drugs*”, “*legal highs*”, “*herbal highs*”, “*bath salts*”, “produtos químicos de pesquisa” e “reagentes”.¹⁶ Historicamente, o termo “*designer drugs*” surgiu em 1984 quando uma grande quantidade de fentanil apareceu no mercado da droga nos Estados Unidos da América (EUA). Mais tarde a comunicação social utilizava o termo “*legal highs*” para falar sobre estes compostos, mas foi no ano de 2005 que a Comunidade Europeia adotou o termo Novas Substâncias Psicoativas, que prevalece até aos dias de hoje.¹⁵

O termo “novas” não se refere necessariamente a substâncias recentes, mas sim substâncias cuja utilização como drogas de abuso é recente, ou seja, o uso de substâncias para fins recreativos.¹⁷ Além disso, muitas destas substâncias já tinham sido sintetizadas anos antes na área farmacêutica com a intenção de desenvolver fármacos, como por exemplo a mefedrona que foi desenvolvida em 1929.¹²

Uma das primeiras NSP a surgir no mercado foi a Ketamina por volta de 1980 na América do Norte, sendo que mais tarde propagou-se em larga escala na Europa e só depois alcançou a Ásia, América do Sul e África Meridional. Na década de 90, surgiram os compostos pertencentes à família das Fenetilaminas e as Piperazinas chegaram ao mercado no início dos anos 2000. Mais tarde, a partir de 2004 surgiram os compostos pertencentes às famílias dos Canabinoides Sintéticos e as Catinonas Sintéticas.¹⁵ Entre 2000 e 2004 era comum a venda de “*party pills*” comprimidos que continham, geralmente, benzilpiperazina e trifluorofenilmetilpiperazina.¹⁸

Em Portugal, o grande crescimento no comércio destas substâncias surgiu em 2007 quando abriu a primeira *smartshop*. Dado a elevada procura destas substâncias, o número destas lojas cresceu de forma exponencial e em 2013 já existiam 63 lojas abertas.^{13,16} No final de 2010, os problemas associados ao consumo das NSP começaram a ser debatidos.¹⁵ Entre 2008 e 2013, o número de apreensões subiu cerca de 7 vezes mais, sendo os compostos pertencentes à família dos Canabinoides Sintéticos vendidos como alternativa ao Cannabis e as Catinonas Sintéticas como substitutos às anfetaminas.¹⁶

A *European Monitoring Centre for Drugs and Drug Addiction* (EMCDDA) lança todos os anos um relatório onde são compiladas todas as informações sobre todas as drogas incluindo,

entre outros, dados estatísticos relativo ao consumo e ao estado do mercado, bem como os compostos emergentes e os efeitos adversos conhecidos. A Figura 1.1 indica de forma esquemática o número de NSP que foram reportadas pela primeira vez em cada ano, de 2005 até 2019.

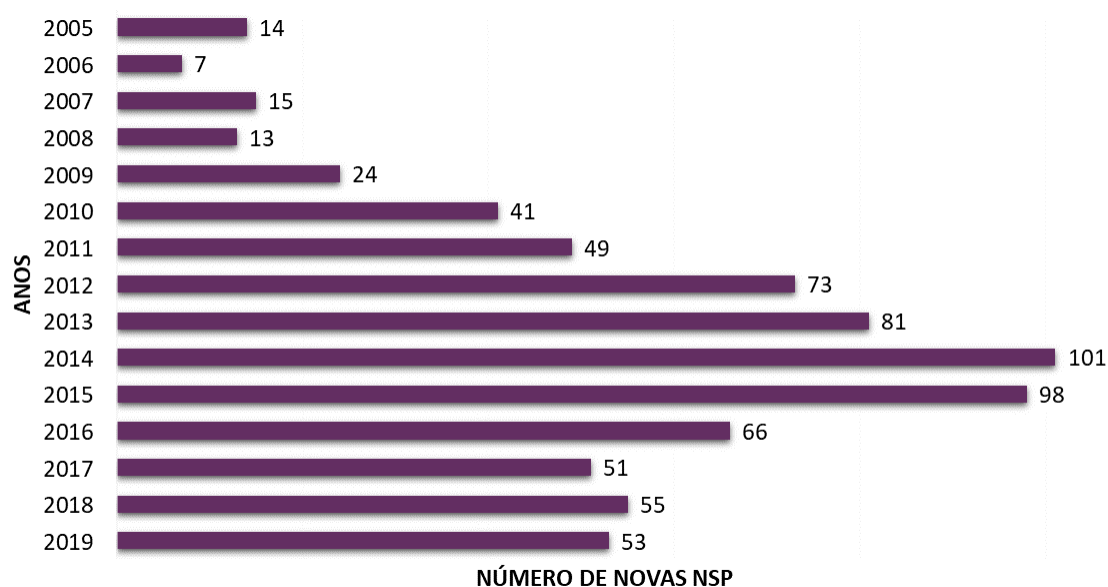


Figura 1.1 Representação gráfica do número de NSP notificadas pela primeira vez em cada ano. Estes dados estatísticos foram retirados dos relatórios anuais sobre drogas entre 2006 e 2020.¹⁹⁻³³

Através desta representação gráfica, podemos ver que foi nos anos de 2014 e 2015 que se registaram um maior número de NSP notificadas pela primeira vez. A implementação das medidas de controlo destas substâncias nos diversos países tem sido apontada como responsável pela diminuição registada nos últimos anos. Na Figura 1.2 estão assinalados alguns marcos importantes na história das NSP nos últimos 15 anos.

Existem dois grandes mercados através dos quais as NSP conseguem chegar à sociedade, *Smartshops* e sites na Internet.¹⁵ Acredita-se que uma das razões do mercado destas substâncias ter crescido de forma exponencial deve-se ao facto de a pureza do ecstasy e da cocaína ter diminuído e, conseqüentemente, a sua procura também.¹⁴ Além disso, o mercado das NSP é bastante ativo e com grande diversidade de substâncias.¹⁸ Um dos maiores produtores é a China que têm apresentado uma evolução constante e variedade de substâncias.¹² O conceito das *Smartshops* surgiu na Holanda no ano de 1994 nas quais eram vendidas substâncias psicoativas que não estavam regulamentadas. Mas foi a partir de 2007 que este conceito se espalhou pela Europa, tornando mais fácil adquirir estas substâncias.¹⁵ Em Portugal, no ano de 2013, foi aprovada a primeira lista de NSP ilícitas, conduzindo ao



Figura 1.2 Alguns marcos importantes na história das NSP entre 2005 e 2020. As informações destacadas neste esquema foram retirados dos Relatórios Anuais sobre droga emitidos pelo EMCDDA entre 2006 e 2020. ¹⁹⁻³³

encerramento forçado de todas as *Smartshops*.¹⁷ Atualmente, é na Internet que os consumidores mais facilmente têm acesso a estas substâncias, em salas de conversação, *blogs*, redes sociais, entre outras plataformas.³⁴ A forma como o mercado apresenta estas substâncias, ou seja, a criatividade e inovação na apresentação das substâncias, atrai os consumidores a comprar através de um simples *click*.¹⁷ Muitas vezes estes compostos são comercializados em embalagens que se assemelham às de produtos lícitos ou com informação enganosa³⁴, usando embalagens coloridas de forma a aliciar e despertar o interesse dos consumidores. Além disso o seu baixo custo é outro dos fatores que incentiva a sua aquisição.¹⁴

Embora o mercado das NSP tenha crescido, o fato de se saber pouco acerca do comportamento que estas substâncias têm no organismo, faz com que os seus efeitos sejam imprevisíveis.³⁴ Por outro lado, o fato de serem substâncias sobre as quais existe pouca informação dificulta também a sua deteção e quantificação, bem como dos seus metabolitos, pois não existem materiais de referência para a sua identificação. Este último ponto apresenta-se como um desafio analítico presente nos laboratórios de Toxicologia Forense.^{14,35} Pelo fato de todos os anos surgirem novas NSP, torna-se difícil o uso de abordagens tradicionais para a sua identificação, por uma lado pela necessidade de existirem materiais de referência, pela dificuldade de detetar estes compostos em pequenas amostras e ainda pela escassez de dados sobre a sua farmacocinética e farmacodinâmica, entre outros.¹⁸

O grupo das NSP representam um conjunto bastante heterogêneo, o que permite dividir os compostos em diferentes famílias.¹⁴ Neste trabalho iremos focar-nos nos Aminoindanos, Canabinoides Sintéticos, Catinonas Sintéticas, Fenetilaminas, Novas Benzodiazepinas, Opióides Sintéticos, Piperazinas, Substâncias tipo Fenciclidinas e Triptaminas.

1.2.1 Aminoindanos

As moléculas presentes na família das Aminoindanos possuem uma estrutura semelhante à da anfetamina devido ao sistema de anel fechado (ver Figura 1.3). A estrutura base pode sofrer diversas

modificações por introdução de grupos funcionais.³⁶ Alguns compostos desta família exercem efeitos psicoestimulantes leves³⁷, contudo caso haja uma overdose no consumo deste composto, os efeitos adversos envolvem taquicardia, hiperpneia, aumento da transpiração,

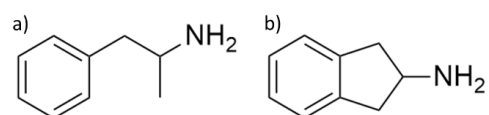


Figura 1.3 Estrutura da anfetamina a) e da 2-aminoindane b).³⁶

desidratação, ansiedade, depressão e paranoia.³⁸ Por um lado, o perfil farmacológico destes compostos pode ser catalogado como inibidores de transporte de monoamina que, de certo modo, desempenha funções na libertação da serotonina.³⁷ Por outro lado, a presença do grupo amina nesta família de moléculas, promove a sua potencial atividade como broncodilatador. Este tipo de moléculas têm sido também estudadas como potenciais analgésicos.³⁹ Relativamente ao seu uso como substâncias de abuso, estudos indicam que em doses recreativas aparentam ser benignos, porém é necessário desenvolver mais estudos para compreender o mecanismo de ação no sistema nervoso central.³⁶

1.2.2 Canabinoides Sintéticos

A síntese de compostos pertencentes à família dos Canabinoides Sintéticos surgiu na década de 70 na tentativa de desenvolver novas moléculas com propriedades anticancerígenas. Assim, os primeiros compostos desta família surgiram em laboratórios académicos e na indústria farmacêutica.⁴⁰ Estas substâncias são sintetizadas de modo a reproduzirem efeitos semelhantes ao do composto (-) - trans- Δ^9 -tetrahydrocannabinol (Δ^9 -THC), a principal substância que reproduz efeitos psicoativos da planta do canábis.⁴¹ De forma semelhante ao Δ^9 -THC, os Canabinoides Sintéticos ligam-se aos recetores de canabinoide tipo 1, CB₁, e tipo 2, CB₂, com diferentes afinidades.⁴² Posteriormente, os laboratórios clandestinos começaram a produção ilícita destes compostos como alternativa à marijuana, incluindo algumas substâncias que não se revelaram promissoras como citostáticos.⁴⁰ Esta família de NSP surgiram no mercado por volta do ano de 2004 como alternativa ao canábis, a um preço mais acessível e com a promessa de não serem detetados em testes utilizados nas medidas de controlo.⁴³ No entanto, foi em 2008 que investigadores forenses na Alemanha e Áustria detetaram o primeiro Canabinoide Sintético, o JWH-018, vendido sob o nome *Spice*. Mais tarde, outros compostos pertencentes a esta família foram encontrados em misturas para fumar e em incensos.⁴¹

Entre todas as famílias das NSP, esta é a que possui o maior número de compostos detetados até à data. Na Figura 1.4 apresenta-se um esquema das estruturas possíveis de Canabinoides Sintéticos. Outro dos aspetos a destacar é o fato de que tem sido reportado um número elevado de intoxicações não fatais associados ao consumo destes compostos, com sintomas que incluem agitação, náuseas e aumento do batimento cardíaco. Em casos mais graves, podem ocorrer convulsões, ataques cardíacos, rutura do tecido muscular, lesão renal,

psicose e vômitos intensos ou prolongados. Adicionalmente, foram também reportados sintomas indicativos de dependência e abstinência.⁴¹

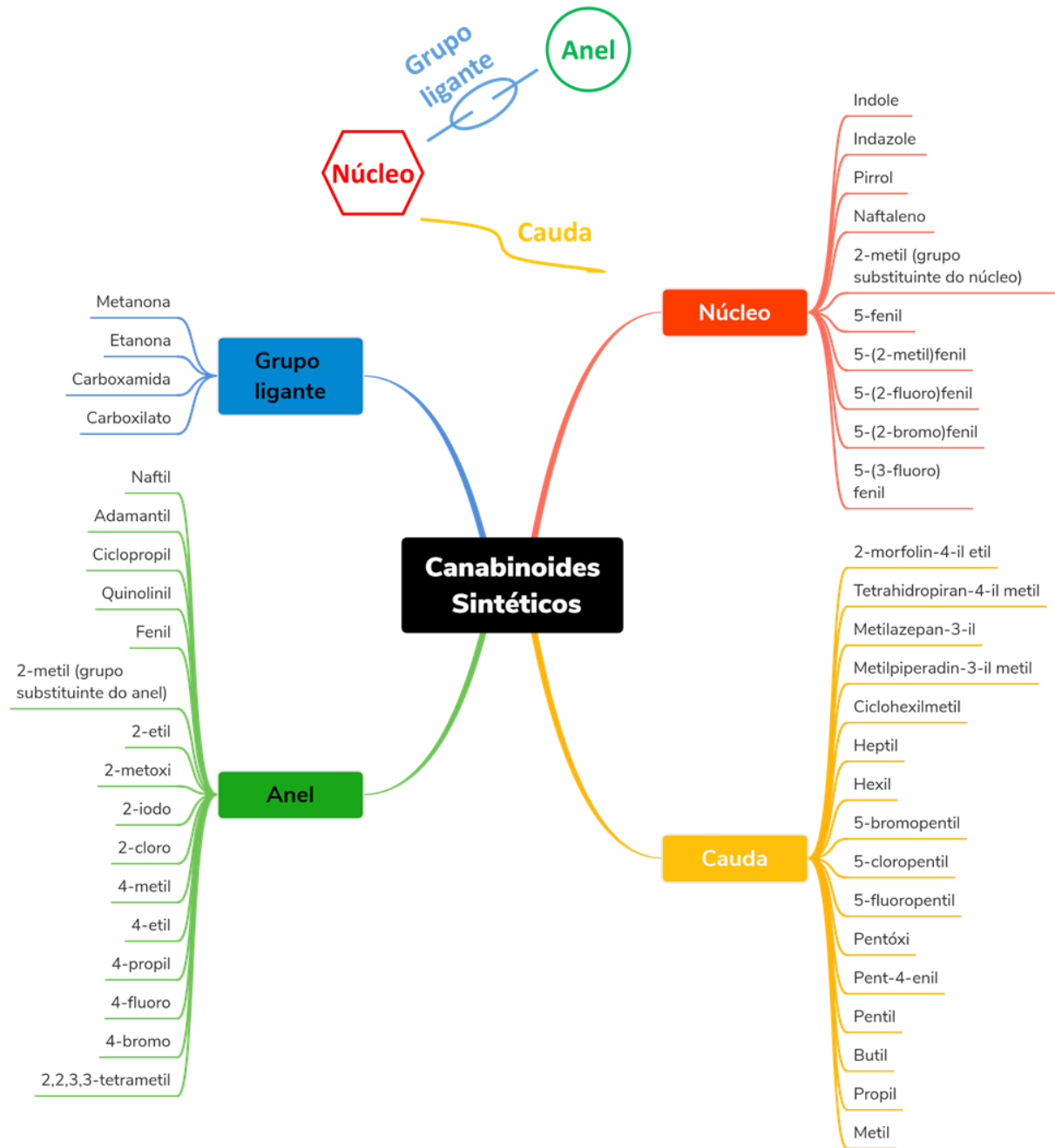


Figura 1.4 Modelo estrutural proposto pelo EMCDDA e os respectivas possibilidades de grupos substituintes para obtenção de diferentes Canabinoides Sintéticos (adaptado de ⁴¹).

1.2.3 Catinonas Sintéticas

As Catinonas Sintéticas são derivadas da (S)-2-amino-1-fenil-1-propanona (catinona) que corresponde a uma substância psicoativa presente na planta *Catha edulis*, mais conhecida por khat.⁴⁴ A síntese das Catinonas Sintéticas surgiu, pela primeira vez, nos anos 20, também como potenciais fármacos. O primeiro composto que surgiu desta família foi a metcatinona e,

de seguida, a mefedrona. A diversidade na estrutura molecular das Catinonas Sintéticas é bastante grande, contudo a principal característica destes compostos é a presença do grupo cetona na posição β da cadeia lateral.⁴⁵

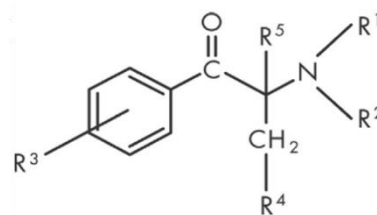


Figura 1.5 Base da estrutura molecular das Catinonas Sintéticas, sendo R^1 , R^2 , R^3 , R^4 e R^5 as possíveis posições para grupos substituintes.⁴⁵

Na Figura 1.5 está representada a estrutura base das Catinonas Sintéticas. Os compostos desta família possuem semelhanças estruturais com as Anfetaminas e atividade biológica semelhante às Fenetilaminas. De entre os compostos estudados até então, apenas o bupropiona possui efeito terapêutico comprovado para o tratamento do transtorno depressivo, em terapias de cessação tabágica e ainda da obesidade.⁴⁴ No caso das Catinonas Sintéticas os principais sintomas adversos reportados são, entre outros, tremores, convulsões, taquicardia, pupilas dilatadas, aumento da temperatura corporal e pressão sanguínea. Os efeitos tóxicos manifestam-se não só no SNC como, também, nos sistemas cardiovascular, gastrointestinal e renal. Além disso, o consumo contínuo destas substâncias pode provocar adição e caso esse consumo seja descontinuado, existe evidência de sintomas de abstinência.⁴⁵

1.2.4 Fenetilaminas

O interesse de compostos das Fenetilaminas no uso recreativo surgiu na década 90 com o lançamento do livro PIHKAL. A venda destas substâncias era, muitas das vezes, realizada nas mesmas redes criminosas que vendiam o Ecstasy. O termo Fenetilamina consiste na contração do seu nome primordial, Feniletilamina.⁴⁶ A estrutura base dos compostos pertencentes a esta família está representada na Figura 1.6. O esqueleto é constituído por um anel aromático ligado a uma cadeia de dois carbonos e um grupo amina no final.⁴⁷ A inserção de determinados grupos substituintes na posição R_6 leva a que o carbono que lhe

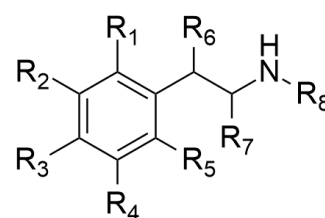


Figura 1.6 Estrutura molecular das Fenetilaminas, sendo que de R^1 a R^8 correspondem às possíveis posições para grupos substituintes.⁴⁷

esteja diretamente ligado passe a ser um centro quiral. Em muitos casos, o comportamento biológico dos estereoisómeros R e S pode ser diferente. Contudo, grande parte dos compostos desta família que são apreendidos encontram-se numa mistura racémica.⁴⁶ Alguns dos sintomas associados ao consumo destas substâncias são sentimentos de angústia e ansiedade, alucinações, taquicardia, tremores e convulsões.⁴⁷

1.2.5 Novas Benzodiazepinas

As Benzodiazepinas são compostos bastante utilizados para o tratamento da ansiedade, epilepsia, espasmo muscular, insônia, entre outros. No entanto, recentemente o uso destas substâncias e dos seus derivados integrou o mercado das drogas ilícitas. Os efeitos provocados com o consumo destas substâncias varia de acordo com a dose, via de administração e eventual mistura com outras drogas, medicamentos ou substâncias. Em doses baixas os efeitos adversos incluem sonolência, fadiga e letargia; no caso de doses mais elevadas, foram reportados distúrbios da coordenação motora, alterações de humor e tonturas.⁴⁸ Estes compostos foram introduzidos como drogas terapêuticas na década de 60. Mais tarde, em 2007 foi detetada a primeira Nova Benzodiazepina com fins recreacionais, a Fenazepam e, posteriormente, Etizolam no ano de 2011.⁴⁹ Relativamente à estrutura molecular, o núcleo destes compostos é constituído por um anel de 7 membros fundido com um benzeno ou tiofeno. O anel de 7 membro pode ter 1 ou 2 átomos de azoto, além disso, normalmente está ligado um fenil ao anel de 7 membros. Na Figura 1.7 apresentam-se algumas estruturas base das Novas Benzodiazepinas.⁵⁰

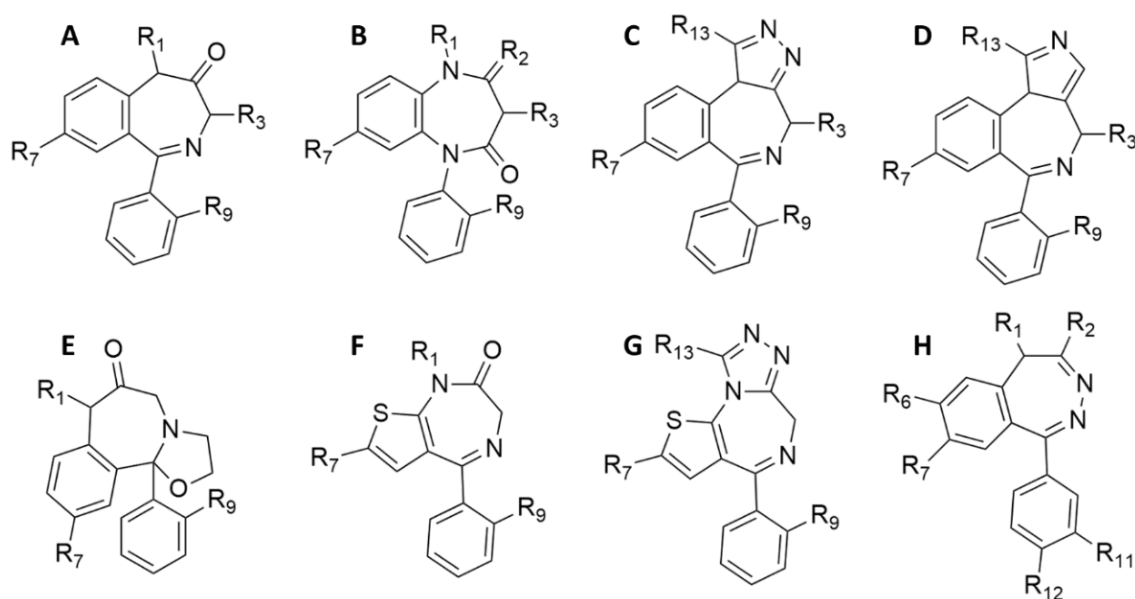


Figura 1.7 Estruturas moleculares base das Novas Benzodiazepinas, sendo que de R¹ a R¹³ correspondem às possíveis posições para grupos substituintes.⁵⁰

1.2.6 Opióides Sintéticos

O termo opióide sintético refere-se a toda e qualquer substância que produza efeitos semelhantes à morfina. Atualmente, a família dos Opióides Sintéticos inclui o fentanil e os seus derivados e os Novos Opióides Sintéticos, ver Figura 1.8.⁵¹ O Fentanil foi sintetizado pela primeira vez no ano de 1960 pelo cientista Paul Janseen na Bélgica e o seu uso foi aprovado para fins médicos, como anestésico. Na década de 80 surgiram casos de uso indevido deste composto. Mais tarde, derivados de Fentanil começaram a ser estudados com o intuito de ser utilizado na área da medicina e veterinária. No final da década de 70 começaram a ser reportados casos de overdose destes compostos, tendo sido por isso, adicionados à lista de NSP a serem controladas. Mais tarde, surgiram os Novos Opióides Sintéticos à venda na Internet, compostos sobre os quais a informação e os estudos são escassos.^{52,53}

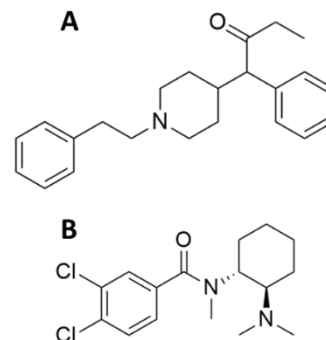


Figura 1.8 Exemplos de dois compostos pertencentes à família dos Opióides Sintéticos: Fentanil (A) e U-47700 (B).⁵²

1.2.7 Piperazinas

Os compostos pertencentes à família das Piperazinas, foram estudados nas décadas 70 e 80 como antidepressivos. Em virtude da sua estrutura molecular, esta família de compostos

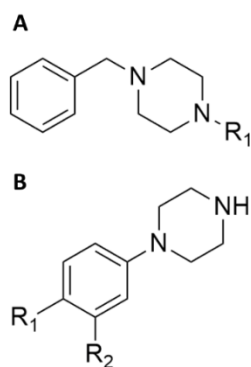


Figura 1.9 Estrutura molecular base das duas classes da família das Piperazinas, derivados de Benzilpiperazina (A) e derivados de Fenilpiperazina (B).⁵⁴

pode ser dividida em duas classes, derivados de Benzilpiperazinas e derivados de Fenilpiperazinas. Em ambas as classes está presente o anel de seis membros com dois átomos de azoto em lados opostos e quatro carbonos entre os dois azotos. Na Figura 1.9 está representada a estrutura base de cada uma destas classes de compostos. O consumo excessivo destes compostos tem sido associado a confusão, taquicardia, ataques epilépticos, vômitos, entre outros.⁵⁴

1.2.8 Substâncias tipo Fenciclidinas

A obtenção do primeiro composto da família das Fenciclidinas, o PCP (Figura 1.10), ocorreu de forma acidental no ano de 1956. Após a sua descoberta, foi explorada a sua potencial aplicação na área da medicina tendo sido também sintetizados compostos análogos com o mesmo propósito. No mercado dos ilícitos os mais comuns são TCP (tenociclidina), PCE (eticiclidina) e PCPy (roliciclidina) e o seu consumo tem sido associado a manifestações de euforia, alucinações, parestesia, entre outros; no caso do consumo de doses elevadas pode ocorrer a dissociação total das sensações e catalepsia, entre outros.⁵⁵

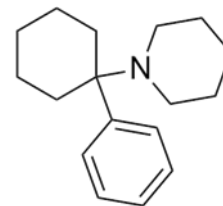


Figura 1.10 Estrutura molecular do PCP.

1.2.9 Triptaminas

Os compostos pertencentes à família das Triptaminas incluem na sua estrutura um grupo indole, ou seja, um anel duplo fundido entre um benzeno e um pirrol como se pode ver na Figura 1.11.⁵⁶ Esta família pode dividir-se em duas classes: Ergolinas, que possuem uma estrutura mais complexa, e as Triptaminas Simples; esta última englobando três subclasses: anel

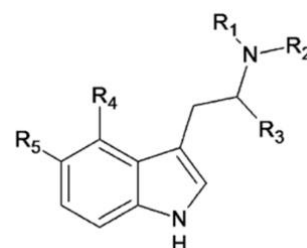


Figura 1.11 Estrutura base das triptaminas.⁵⁷

não substituído, anel substituído na posição 4 e anel substituído na posição 5 (Figura 1.12). O consumo destes compostos está associado a ansiedade, vômitos, dor abdominal, midríase, taquicardia, taquipneia, hipertensão, entre outros, e hipertermia em caso de consumo excessivo. O fato destes sintomas se manifestarem por curtos períodos de tempo tem sido apontado como um risco de falsa segurança para o seu consumo⁵⁷, induzido ao consumo de doses elevadas que por vezes resultam em morte.⁵⁶

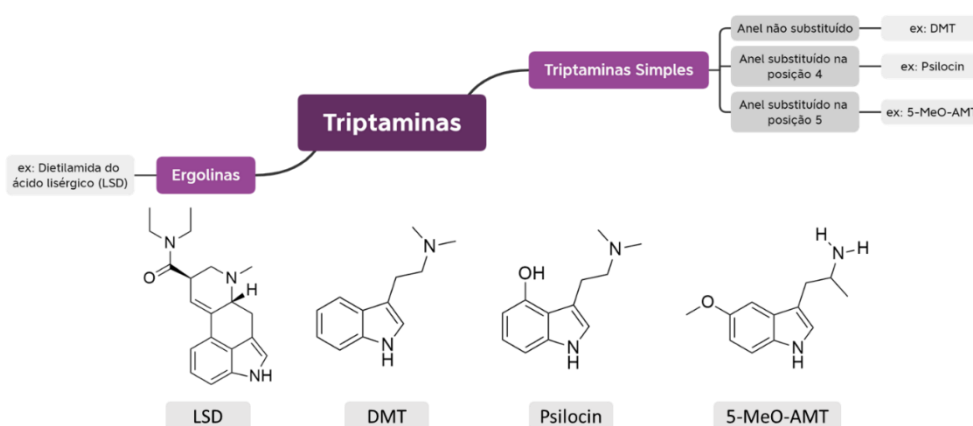


Figura 1.12 Classificação das Triptaminas (adaptado de ⁵⁶).

1.3 Objetivos

Este projeto pretende contribuir para uma melhor classificação e caracterização das NSP desconhecidas sobre as quais não existe ainda regulamentação. Para tal, recorrer-se-á a dois métodos de análise não supervisionada, *Principal Component Analysis*, PCA e *Hierarchical Cluster Analysis*, HCA. O PCA permite selecionar o conjunto de descritores moleculares para a discriminação entre famílias de NSP. Com o método HCA, avaliar-se-á a similaridade estrutural das NSP e respetivas famílias. A ideia principal prende-se com criação de uma metodologia que permita caracterizar e classificar de forma rápida e eficiente NSP desconhecidas tendo por base o perfil químico de NSP já regulamentadas. Para explicar os agrupamentos e os perfis químicos identificados com os métodos de aprendizagem computacional, serão selecionadas NSP para uma caracterização mais detalhada recorrendo à simulação molecular, por Dinâmica Molecular e *Docking*. A Dinâmica Molecular permitirá avaliar a interação entre as NSP e uma membrana modelo. O *Docking* permite analisar as interações entre as NSP e o recetor-alvo.

1.4 Aplicação de métodos computacionais na caracterização de drogas

O uso de técnicas computacionais têm crescido nas últimas décadas, o que levou ao desenvolvimento de métodos sofisticados para resolver problemas complexos. Independentemente da importância que os métodos tradicionais têm, estes demoram muito tempo e os seus custos são muitas vezes elevados.⁵⁸ Neste sentido, o uso de técnicas computacionais permitem dar suporte às técnicas consideradas mais tradicionais, uma vez que computacionalmente é possível analisar uma grande quantidade de resultados em pouco tempo.⁵⁹ Por outro lado, a sua utilização permite utilizar uma vasta quantidade de ferramentas, incluindo bases de dados, *softwares* que permitem a descrição das moléculas por descritores moleculares, simuladores de sistemas biológicos e dinâmica molecular, entre outros.⁵⁸

Na categoria das NSP a maioria dos trabalhos desenvolvidos têm-se centrado essencialmente nos Canabinoides Sintéticos e nas Catinonas Sintéticas, sendo, a família das Substâncias tipo Fenciclidinas a menos estudada de todas.^{41,43,52}

Os estudos envolvem maioritariamente a combinação de métodos computacionais e técnicas experimentais. Por exemplo, Risoluti *et al.*⁶⁰, recorreu à Espectroscopia de Infravermelhos e aos Métodos Quimiométricos, nomeadamente o PCA, para elaborar uma

metodologia para identificar Canabinoides Sintéticos e Fenetilaminas em amostras apreendidas.⁶⁰ Uma metodologia semelhante foi utilizada por Omar *et al.*⁶¹ recorrendo a dados obtidos por Espetroscopia de Raman.⁶¹ Para demonstrar que esta metodologia seria igualmente eficaz para espectrómetros portáteis, Metternich *et al.*⁶² procurou validar esta metodologia utilizando a Espetroscopia de Raman portátil e comparou os resultados com os obtidos com a Espetroscopia de Raman clássica. Os resultados obtidos com ambas as técnicas foram bastante satisfatórios o que demonstrou ser vantajoso para análises analíticas fora do laboratório, como é o caso nos controlos de fronteiras.⁶² Num outro trabalho, Fowble *et al.*⁶³, combinaram os dados analíticos das Catinonas Sintéticas obtidos através da Análise Direta em Tempo Real (DART, do inglês *Direct Analysis in Real Time*) acoplada à Espectrometria de Massa de Alta Resolução (HRMS, do inglês *High Resolution Mass Spectrometry*) com HCA e PCA. Usando como dados os valores m/z de fragmentos perdidos ao longo da fragmentação, os autores utilizaram o PCA para discriminar as possíveis subclasses pertencentes à família das Catinonas Sintéticas e, de seguida, o HCA para avaliar a similaridade das estruturas moleculares. Com esta metodologia, conseguiram agrupar as moléculas da família das Catinonas Sintéticas em subclasses e, também, determinar com sucesso a que grupo pertenciam amostras desconhecidas, tendo estabelecido um método de triagem eficiente para laboratórios forenses.⁶³

No que se refere à aplicação da simulação computacional, alguns trabalhos sobretudo usando a Dinâmica Molecular (DM), têm sido reportados com o intuito de relacionar as NSP com a sua atividade biológica. Gaspar e os seus colaboradores⁶⁴ procuraram avaliar a toxicidade de novas Catinonas Sintéticas, caracterizar compostos estruturalmente semelhantes e comparar o seu comportamento biológico, recorrendo à Dinâmica Molecular, em sistemas contendo cada um dos compostos em interface numa membrana modelo de 1-palmitoil-2-oleoil-sn-glicero-3-fosfolina, POPC. Os resultados demonstraram que, apesar da elevada similaridade molecular, o comportamento de cada substância ao atravessar a membrana era diferente, sendo por isso espectável que a biodisponibilidade de cada composto varie substancialmente. Esta propriedade estará também eventualmente correlacionada com a toxicidade celular. Adicionalmente, observaram que as Catinonas Sintéticas em estudo sofreram uma desprotonação para atravessar a bicamada lipídica, isto acontece porque estas moléculas adaptam-se ao ambiente para que a sua mobilização seja facilitada.⁶⁴

Também recorrendo à simulação por DM, Ametovski et al.⁶⁵ comparou a interação de compostos da família dos Canabinoides Sintéticos com os recetores CB₁ e CB₂. Os resultados mostraram a afinidade é superior, na sua generalidade, para derivados CUMYL em ambos os recetores. Por outro lado, os derivados benzílicos são os que possuem menor afinidade enquanto os derivados de α -metilbenzilamina apresentam uma afinidade intermédia. Além disso, os derivados de α -metilbenzilamina demonstraram ser agonistas potentes e eficazes de CB₁, já para o recetor CB₂ os valores de eficácia foram baixos, tal como os derivados benzílicos.⁶⁵ Recentemente, Vo et al.⁶⁶ utilizaram a Dinâmica Molecular para explorar a interação entre o fentanil e o recetor μ -opioide, uma vez que grande parte das mortes ocorridas nos Estados Unidos por overdose estão relacionadas com o consumo de opioides sintéticos. Os resultados mostraram que para além da ligação por intercalação de sal comum aos opioides derivados da morfina, o fentanil se liga ao recetor através de uma ligação de hidrogénio estável com a His297. Esta ligação adicional apenas está disponível quando a His297 adota o tautómero neutro. Segundo os autores, estes resultados, fornecem uma base para investigações farmacológicas de análogos do fentanil, particularmente de como as modificações estruturais alteram as propriedades de ligação dos derivados do fentanil, que podem ter um potencial de abuso acrescido.⁶⁶

Capítulo 2

Teoria e Métodos

2. Teoria e Métodos

Neste Capítulo será apresentado o enquadramento teórico das técnicas quimiométricas e dos métodos de simulação molecular que foram usadas ao longo deste trabalho.

2.1 Quimiometria e aprendizagem computacional aplicado à Química Forense

O termo Quimiometria surgiu no início da década de 70 com os cientistas Svante Wold e Bruce R. Kowalski⁶⁷ para descrever a disciplina da química que analisa e interpreta conjuntos de dados químicos utilizando métodos matemáticos e estatísticos.⁶⁸ Pouco tempo depois da introdução desta disciplina surgiu a Sociedade Internacional da Quimiometria⁶⁷ e não tardou para começarem a ser publicados artigos nesta área, tendo surgido a primeira publicação em 1975. Mais tarde, foram criadas duas revistas científicas dedicadas ao estudo desta disciplina: *Chemometrics and Intelligent Laboratory Systems* e *Journal of Chemometrics*.⁶⁹

Até à data, várias foram as definições propostas, contudo a definição atual é dada por:

Quimiometria → disciplina da área da química que utiliza métodos matemáticos e estatísticos para desenvolver e otimizar metodologias para promover o máximo de informação química sobre o conjunto de dados.⁶⁷

A Quimiometria é amplamente utilizada nas diversas sub-disciplinas da Química (como é o caso da Química Analítica e Química Forense, entre outras), mas também fora desta área, como é o caso da Biologia.⁷⁰

As principais vantagens da sua utilização são a capacidade de analisar, interpretar e visualizar um grande conjunto de dados descritos por várias variáveis ao mesmo tempo, permitindo avaliar a correlação entre essas variáveis, bem como a complementaridade com as metodologias experimentais.⁷¹

Neste estudo foi implementado o método da Análise de Componentes Principais (PCA) para determinar os descritores moleculares mais relevantes para a discriminação entre famílias de NSP. O método da Análise Hierárquica de Agrupamentos (HCA) permitiu avaliar a similaridade estrutural das NSP e respetivas famílias, tendo por base os descritores moleculares selecionados no método PCA. Com estes métodos quimiométricos pretendia-se

caracterizar e classificar de forma rápida e eficiente as NSP desconhecidas e não regulamentadas, tendo por base o perfil químico das estruturas das NSP já controladas.

Uma vez agrupadas as NSP foram utilizados métodos de simulação computacional, nomeadamente a *Dinâmica Molecular* e o *Docking Molecular* por forma a avaliar se a semelhança/dissemelhança obtida pelo HCA se reflete no comportamento de algumas moléculas selecionadas na presença de um recetor biológico e de uma membrana modelo. Com o método de Dinâmica Molecular pretendia-se estudar o comportamento de diferentes NSP numa membrana modelo; com método *Docking* molecular pretendia-se avaliar a interação entre um recetor e diferentes NSP. Resumidamente, pretendeu-se avaliar se moléculas com perfis químicos semelhantes têm comportamento biológicos semelhantes. A finalidade deste estudo é, então, obter uma metodologia que nos permita caracterizar e classificar NSP desconhecidas e não regulamentadas.

Atualmente, os métodos de aprendizagem computacional (do inglês, *machine learning*) têm sido aplicados em diversas áreas da química. No campo das NSP, alguns exemplos incluem o trabalho de Pereira *et al*, que usou o método de PLS-DA (*Partial Least Squares – Discriminant Analysis*) em dados espectrais de Catinonas Sintéticas e Triptaminas recolhidos por ATR-FTIR.⁷² Outro trabalho centrou-se nas famílias dos Canabinoides Sintéticos e Catinonas Sintéticas, por forma a obter o perfil farmacológico dos compostos em estudo através do uso do PCA e modelos de classificação.⁷³

As estratégias de aprendizagem computacional aplicadas a sistemas deste tipo dependem de critérios impostos (designando-se por técnicas supervisionada) ou são definidas de acordo com a similaridade dos padrões (técnicas não supervisionadas), independente de critérios impostos. Estas técnicas permitem simplificar estruturalmente os dados multivariados, identificar quais as variáveis/descriptores mais característicos e diferenciadores, e caracterizar tendências. Neste estudo, foram selecionados dois métodos não supervisionados, a Análise Hierárquica de Agrupamentos (em inglês *Hierarchical Cluster Analysis, HCA*) e a Análise de Componentes Principais (em inglês *Principal Component Analysis, PCA*).

O primeiro método de análise é usado neste estudo para definir a estrutura natural dos dados, associando as NSP com características estruturais comuns. A associação sucessiva entre NPS é traduzida numa estrutura hierárquica, designada por dendrograma, na qual as NSP semelhantes, segundo os parâmetros escolhidos, são agrupados entre si. Por outro lado,

a Análise de Componentes Principais é usada para reduzir a dimensionalidade do conjunto de dados, descrito por um conjunto de descritores moleculares inter-relacionados associados à estrutura, conformação e topologia das NSP, mantendo tanto quanto possível a variabilidade presente no conjunto de dados original.

2.2 Análise Hierárquica de Agrupamentos

A Análise Hierárquica de Agrupamentos (HCA) é um método que, num conjunto de objetos (neste caso as NSP) aparentemente heterogêneos, procura identificar subgrupos que apresentam homogeneidade entre si e heterogeneidade com outros grupos através de medidas de similaridade. Uma das principais características deste método é o fato de incorporar um processo não supervisionado, ou seja, sem se basear em grupos inicialmente definidos nem conhecer a informação individual de cada NSP.⁷⁴ Este processo de formação de subgrupos pode decorrer através de duas metodologias diferentes: o método aglomerativo ou o método divisivo.⁷⁵ No primeiro método, as moléculas numa fase inicial encontram-se separadas umas das outras e vão agrupando-se em etapas. No método divisivo, as moléculas encontram-se num único grupo e, de seguida, separam-se em dois grupos, e assim sucessivamente até que já não seja possível mais divisões de grupos.⁷⁴

Independentemente se é utilizado o método aglomerativo ou o divisivo, a representação gráfica mais comum é o dendrograma. Na Figura 2.1 é possível observar um exemplo de um dendrograma, os objetos numerados de 1 a 5 estão ligados por linhas, esta ligação foi tendo por base a similaridade entre objetos do mesmo grupo e dissimilaridade entre grupos.

Para obtermos o dendrograma final, é preciso realizar alguns passos. O primeiro passo envolve preparar o conjunto de dados, ou seja, o conjunto de dados tem de estar organizado numa matriz numérica onde as linhas representam os objetos (moléculas) e as colunas as variáveis (descritores moleculares). Para terminar este passo, é necessário normalizar os dados para que a comparação de descritores seja possível, pois provavelmente possuem escalas diferentes. De seguida, é necessário escolher qual será a distância utilizada para unir moléculas e grupos, com o intuito de construir a matriz de distâncias.⁷⁶ Existem inúmeras distâncias que podem ser utilizadas para o efeito, como por exemplo a distância Euclidiana e

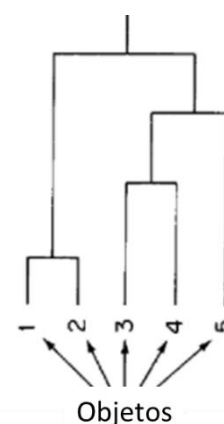


Figura 2.1 Exemplo de uma representação gráfica obtida pelo HCA, o dendrograma.

a distância de Manhattan, sendo que as expressões estão apresentadas nas equações 2.1 e 2.2, respetivamente.⁷⁷


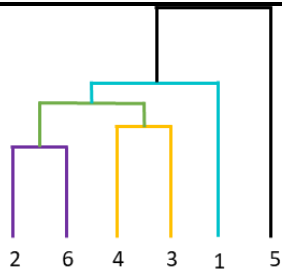
$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{in} - x_{jn}|^2} \quad (2.1)$$

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}| \quad (2.2)$$

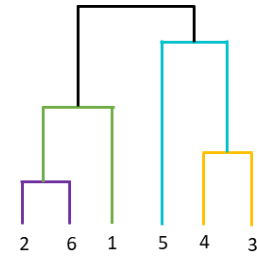
onde x_i e x_j correspondem aos objetos em que a distância está a ser medida nas n dimensões. O terceiro passo consiste em escolher o critério de ligação mais apropriado. Os critérios de ligação mais conhecidos são a ligação simples (em inglês *single linkage*), ligação completa (em inglês *complete linkage*), ligação média (em inglês *average linkage*) e, também, o método *Ward*.⁷⁸

No caso da ligação simples a distância entre dois grupos é definida como a distância entre os seus dois objetos mais próximos. Geralmente, é visível no dendrograma que os objetos foram adicionados sequencialmente ao grupo. Já a ligação completa define a distância entre dois grupos como a distância entre os seus dois objetos mais afastados. O dendrograma aplicando este critério tende a ser bem definido e compacto. O critério da ligação média define a distância entre dois grupos como a distância média entre os pares de objetos, de modo que todos os objetos dos dois grupos tenham a mesma influência no resultado final. Por fim, o critério *Ward* tem por base a análise da variância, agrupando os objetos de forma a que esta variância seja mínima.⁷⁸ Na tabela 2.1 é possível visualizar como é feita a associação e exemplos de dendrogramas utilizando os diferentes critérios de ligação.

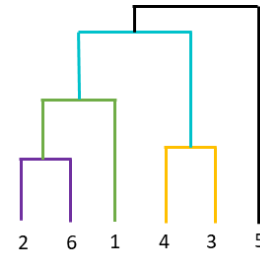
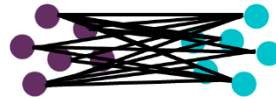
Tabela 2.1 Sumário dos critérios de ligação mais conhecidos e o respetivo dendrograma.

Critério de Ligação	Associação	Dendrograma
Ligação Simples (<i>single linkage</i>)		

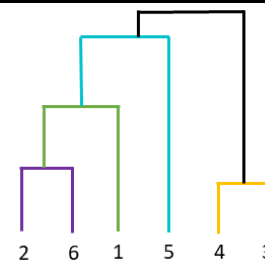
Ligação Completa
(*complete linkage*)



Ligação Média
(*average linkage*)



Ward



Tendo, então, escolhido a distância e o critério de ligação é possível reproduzir a análise HCA e obter o dendrograma do conjunto de dados. Assim sendo, o último passo consiste em interpretar a representação gráfica obtida.

2.3 Análise de Componentes Principais

A Análise de Componentes Principais (PCA) é o método de análise não supervisionada mais utilizado para redução de dimensionalidade, permitindo analisar conjuntos de dados de grandes dimensões e de elevada complexidade.⁷⁹

Especificamente, o PCA permite reduzir a dimensionalidade do conjunto de dados que possui um elevado número de variáveis/descriptores correlacionadas, enquanto mantém a maior percentagem possível de informação estatística. A redução é feita através da transformação das variáveis originais num novo conjunto de variáveis e são designadas por Componentes Principais (em inglês *Principal Components*, PC's). Estas novas variáveis não estão correlacionadas e são, preferencialmente, em menor número quando comparado com o conjunto de variáveis original.⁸⁰ Quando se diz que este método reduz o número de variáveis

preservando o máximo de variabilidade possível, na verdade o que o método faz é encontrar novas variáveis que são funções lineares das variáveis iniciais. Depois, são escolhidas as funções lineares com variância máxima, visto que as componentes principais obtidas devem possuir na medida do possível o máximo de informação dos dados.⁸¹ Portanto, a redução do número de variáveis irá tornar o estudo mais adequado e eficiente ao analista. Na análise às componentes principais, será possível confirmar que a informação relevante se encontra presente nas primeiras componentes principais e que, por consequência, as restantes não são tão significativas pois a sua variância é pequena. Após a escolha do número de componentes principais é possível analisar os resultados através de uma representação gráfica, normalmente em duas dimensões. Esta análise irá permitir que o analista confirme ligações de informação já conhecidas, mas, também, novas relações para as quais não havia quaisquer indícios.⁸⁰

2.3.1 Passos na aplicação do PCA

A aplicação do método do PCA implica um conjunto de etapas que irão ser descritas a seguir.

1. Escolha do software a ser utilizado para processar as funções necessárias para a aplicação do método. Neste estudo, foi utilizado o software R e o seu ambiente de desenvolvimento integrado (IDE), o RStudio.
2. Seleção das variáveis que irão ser alvo de estudo. Para este projeto as variáveis utilizadas referem-se a descritores presentes no pacote CDK (*Chemistry Development Kit*) do R.
3. Escolha da matriz a processar, as possibilidades são entre a matriz de Covariância e a matriz de Correlação. No caso da primeira, a escolha acontece quando o conjunto de variáveis inicial detém da mesma unidade de medida e variâncias próximas. Por outro lado, a matriz de Correlação é escolhida quando o conjunto de inicial de variáveis não possui a mesma unidade de medida ou caso possuam variância muito diferenciada.
4. Definir o número de componentes principais mais indicado para a análise, para tal existem três critérios: Critério de Pearson, Critério de Kaiser e Scree Plot.⁸⁰
5. O passo final consiste na análise e interpretação dos resultados obtidos, como por exemplo da representação gráfica que muitas das vezes surge como biplot. Um biplot é um gráfico de dispersão, com duas ou três dimensões, que permite representar simultaneamente as variáveis/descriptores e os objetos/moléculas.

Obtenção das componentes principais através da matriz de Covariância ou Correlação

Assumindo $X = [X_1 \dots X_p]'$ um vetor aleatório que possui um valor médio μ e uma matriz de variância/covariância Σ ,

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix} \quad (2.3)$$

Onde,

$$\sigma_{ij} = Cov(X_i, X_j) \quad (2.4)$$

Admitindo que $X_{(n \times p)}$ corresponde a uma matriz de dados com dimensão n , de observações deste vetor aleatório:

$$X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{p1} & \cdots & X_{pp} \end{bmatrix} \quad (2.5)$$

O produto obtido corresponde a uma rotação do sistema ortogonal de eixos associados às variáveis iniciais (matriz $X_{(n \times p)}$), em que, após a sua aplicação, é obtido um novo sistema de dados (matriz $Y_{(n \times p)}$). As colunas da matriz Y são as novas variáveis Y_j , as componentes principais. O que se pretendia com este passo era encontrar um novo conjunto de p variáveis (componentes principais, $Y_1 \dots Y_p$) não correlacionadas e que apresentem a máxima variância.

As componentes principais correspondem a combinações lineares das p variáveis da matriz X :

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{pj}X_p = a'_jX, j = 1, \dots, p \quad (2.6)$$

onde $j = 1, \dots, p$ e a_{ij} ($i = 1, \dots, p; j = 1, \dots, p$) são constantes.

De modo a determinar os coeficientes destas combinações lineares, existe um conjunto de condições que têm que ser cumpridas:

1. $Var(Y_1) \geq Var(Y_2) \geq \dots \geq Var(Y_p)$
2. As componentes principais não são correlacionadas entre si.
3. Independentemente da componente principal, a soma dos quadrados dos coeficientes é igual a 1 ($Y_i: a_{1j}^2 + a_{2j}^2 + \dots + a_{pj}^2 = 1$).

Do conjunto de condições mencionadas, podemos concluir que Y_1 é a componente principal que possui maior variância; Y_2 é a segunda componente com maior variância e não está correlacionada com Y_1 ; Y_3 é a terceira componente com maior variância e não está correlacionada com Y_1 e Y_2 ; e assim continuamente.

Deste modo, na equação 2.6 (a_1, a_2, \dots, a_p) diz respeito aos p vetores próprios associados aos p maiores valores próprios de Σ ($\lambda_1 > \lambda_2 > \dots > \lambda_p$) e $Var(Y_j) = \lambda_j$. Dado que as componentes principais foram determinadas com o intuito de não serem correlacionadas, duas a duas, conseqüentemente, teremos uma covariância nula:

$$Cov(Y_j, Y_{j'}) = a_j' \Sigma a_{j'} = a_j' \lambda_j a_{j'} = \lambda_j a_j' a_{j'} = 0 \quad (2.7)$$

ou seja, $a_j' a_{j'} = 0$, o que quer dizer que a_j' e $a_{j'}$ (com $j \neq j'$) são vetores ortogonais.

No caso em que as variáveis originais não possuem a mesma unidade, escala, ou até mesmo natureza distinta, é necessário uniformizar o conjunto de dados. Para alcançar esta uniformização, divide-se cada valor pelo desvio padrão da variável centrada correspondente. Conseqüentemente, todas as variáveis passam, assim, a ter todas a mesma variância unitária e valor médio nulo. A matriz de covariância do *novo* conjunto de variáveis é igual à matriz de correlação do conjunto das variáveis iniciais, visto que:

$$Cov\left(\frac{X_i}{\sigma_i}, \frac{X_j}{\sigma_j}\right) = \frac{Cov(X_i, X_j)}{\sigma_i \sigma_j} = Corr(X_i, X_j) \quad (2.8)$$

Para um conjunto de dados deste tipo, a análise é realizada utilizando a matriz de correlação, P . À semelhança do processo anterior, a obtenção das componentes principais é feita através dos valores e vetores próprios da matriz de correlação, que neste caso é denominada por matriz P . É evidente que os vetores próprios de P não são iguais aos da matriz Σ , o mesmo acontece com as componentes principais. A matriz P define-se como uma matriz simétrica de diagonal com valor 1:

$$P = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{bmatrix} \quad (2.9)$$

onde,

$$\rho_{ij} = Corr(X_i, X_j) \quad (2.10)$$

Redução da dimensionalidade

Como já foi referido, o principal objetivo do método PCA é reduzir a dimensionalidade do conjunto de dados. Para tal, é criado um novo conjunto de variáveis não correlacionadas a partir das variáveis originais. A Figura 2.2 representa, em forma esquemática, o que realmente acontece aquando de uma redução da dimensionalidade.

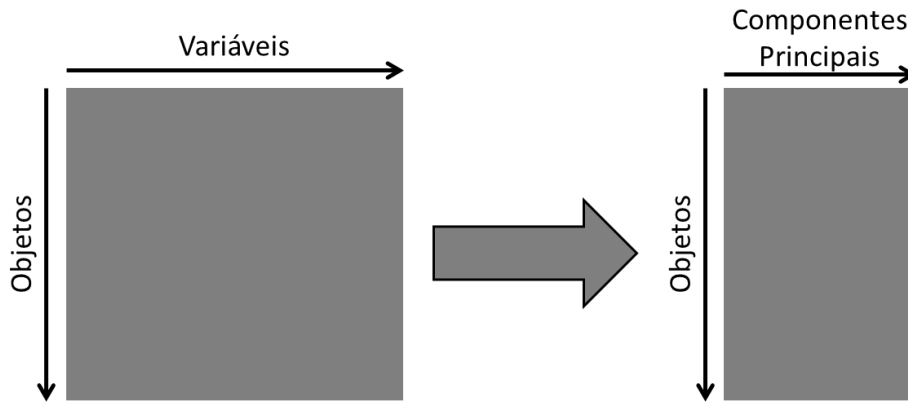


Figura 2.2 Representação esquemática do processo de redução da dimensionalidade.

Para realmente reduzir a dimensionalidade é necessário escolher quais as componentes principais que se devem considerar. É possível ordenar as componentes principais por ordem decrescente da sua variância e quanto maior for o seu valor, mais informação relativamente ao conjunto inicial tem. A variância de cada componente principal é dada por:

$$Var(Y_j) = \lambda_j \quad (2.11)$$

Portanto, a soma das variâncias das componentes principais é dada pela expressão:

$$\sum_{j=1}^p Var(Y_j) = \sum_{j=1}^p \lambda_j \quad (2.12)$$

No caso da matriz Σ , que é uma matriz simétrica, a soma dos seus valores próprios é igual ao traço da matriz, ou seja, ao produtos dos elementos da diagonal principal, logo:

$$tr(\Sigma) = \sum_{j=1}^p Var(X_j) \Rightarrow \sum_{j=1}^p \lambda_j = \sum_{j=1}^p Var(X_j) \quad (2.13)$$

e assim,

$$\sum_{j=1}^p Var(X_j) = \sum_{j=1}^p Var(Y_j) \quad (2.14)$$

Com a expressão 2.14 é possível afirmar que a soma das variâncias das variáveis originais é igual à soma das variâncias das componentes principais e que se forem analisadas o conjunto total das componentes principais é possível explicar toda a variabilidade. Com isto, a proporção da variância total que é explicada pela j -ésima componente principal Y_j e que indica a importância da mesma é dada por:

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{tr(\Sigma)} \quad (2.15)$$

Para o caso da matriz de correlação, a variância total será igual ao número de variáveis p

$$\sum_{j=1}^p \lambda_j = \sum_{j=1}^p \text{Var}(Y_j) = \sum_{j=1}^p \text{Var}(X_j) = p \quad (2.16)$$

Do mesmo modo, a proporção da variância total que é explicada pela j -ésima componente principal Y_j e que indica a importância da mesma é dada por:

$$\frac{\lambda_j}{\sum_{j=1}^p \lambda_j} = \frac{\lambda_j}{p} \quad (2.17)$$

Para determinar o número de componentes ideal para cada estudo, existem vários critérios que podem ser utilizados, nomeadamente o Critério de Pearson (ou regra dos 80%), Critério de Kaiser e *Scree plot*.

(i) Critério de Pearson (ou regra dos 80%)⁸⁰

O Critério de Pearson tanto pode ser utilizado em matrizes de Covariância, como em matrizes de Correlação. O número de componentes principais escolhidas têm de reter pelo menos 80% da informação total ou da variabilidade total. Depois de as ordenarmos da maior para a menor variância, devem considerar-se tantas componentes principais quantas necessárias para que a percentagem de variância seja superior a 80%, ou seja, conservar as primeiras r componentes principais de modo que

$$\frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.80 \quad (2.18)$$

(ii) Critério de Kaiser⁸⁰

O Critério de Kaiser é apenas utilizado no caso da matriz de Correlação. Neste critério as componentes são escolhidas pelo seu valor próprio e, por isso, são consideradas apenas as componentes que possuem um valor próprio superior à unidade (o valor unitário corresponde à média do conjunto de valores próprios).

$$\bar{\lambda} = \frac{1}{p} \sum_{j=1}^p \lambda_j \quad (2.19)$$

(iii) Scree Plot⁸⁰

Finalmente, o terceiro critério Scree Plot possibilita a determinação do número de componentes principais através da visualização de uma representação gráfica (Figura 2.3). A representação gráfica surge em modo de gráfico onde são representados os pontos de abcissa j e a ordenada igual j -ésimo valor próprio ou à percentagem de variância explicada pela j -ésima componente principal, isto é, pontos de coordenadas (j, λ_j) ou $(j, \lambda_j / \sum_{j=1}^p \lambda_j)$ onde se diferenciam as contribuições das diversas componentes principais. De acordo com este critério, são escolhidas as que mais contribuem, sendo que estas se encontram mais destacadas em relação às restantes na representação gráfica.

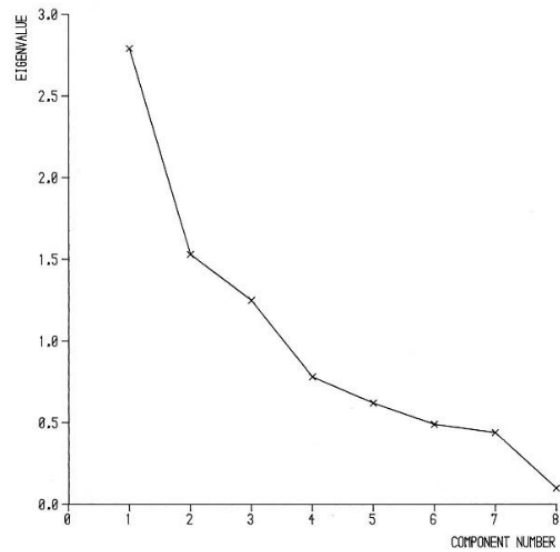


Figura 2.3 Representação gráfica dos valores próprios em função do número de componentes consideradas (adaptado de ⁷⁸).

Scores e Loadings

Designa-se por *Scores* as coordenadas das amostras no novo sistema de referência. Como já foi referido no ponto anterior, as componentes principais são o resultado de uma transformação (combinação linear) sobre as variáveis em estudo. A mesma transformação pode ser efetuada para os dados, quer dizer, aos vetores de observações x_1, x_2, \dots, x_p (colunas da matriz de dados X) das variáveis X_1, X_2, \dots, X_p , respetivamente. Com isto, teremos uma nova matriz de dados, a matriz Y , com dimensão $(n \times p)$ em que o ij -ésimo elemento terá um valor igual ao *score* do i -ésimo para a j -ésima componente principal:

$$y_{ij} = a_{1j}x_{i1} + a_{2j}x_{i2} + \dots + a_{pj}x_{ip} \tag{2.20}$$

Portanto, a matriz dos *scores* dos objetos será dela pela matriz:

$$Y = \begin{bmatrix} y_{11} & \dots & y_{1r} \\ \vdots & \ddots & \vdots \\ y_{n1} & \dots & y_{nr} \end{bmatrix} \tag{2.21}$$

As *Loadings*, por outro lado, representam os coeficientes da combinação linear que descreve cada componente principal, ou melhor, o impacto de cada variável original em cada componente principal.

2.4 Modelação e Simulação Molecular

As simulações e as abordagens de aprendizagem computacional têm sido amplamente utilizadas, por exemplo, no processo de desenvolvimento de novos fármacos, uma vez que fornecem informações, a nível molecular, do comportamento de uma dada molécula no meio biológico, na presença de recetores, membranas e/ou de outras moléculas.⁸² Dentro destas abordagens computacionais destacam-se o *Docking* e a Dinâmica Molecular (DM). O objetivo da modelação computacional de macromoléculas, tais como as proteínas (também designadas por recetores) é fornecer uma alternativa à utilização das técnicas analíticas de Cristalografia e Ressonância Magnética Nuclear (RMN), e mimetizar a estrutura tridimensional de uma macromolécula de interesse e cuja estrutura não se encontre resolvida experimentalmente.

Neste contexto, o *Docking* e a DM, permitem avaliar o comportamento ao nível molecular dos sistemas envolvendo macromoléculas e moléculas pequenas, designadas de ligandos. A estabilidade dos complexos formados entre a macromolécula e o ligando é governada por interações intermoleculares, isto é, interações não-covalentes. As interações mais comuns neste tipo de sistemas são as ligações de hidrogénio, as interações hidrofóbicas, eletrostáticas e de van der Waals.^{84,86}

Nos últimos 20 anos, mais de 60 ferramentas e programas de *Docking* foram desenvolvidos para utilização académico e comercial, como o DOCK, AutoDock, FlexX, LigandFit, AutoDock Vina, entre muitos outros.⁸⁵ Estes programas diferem essencialmente nas estratégias de posicionamento do ligando e são classificados como algoritmos (i) de construção incrementais, como FlexX, (ii) baseados na forma (ex. DOCK), (iii) genéticos (GOLD), (iv) de pesquisa sistemática (Glide, Schrödinger) e (v) baseados em simulações de Monte Carlo (LigandFit). Com exceção do GOLD, quase todos os programas atuais de *Docking* de ligandos flexíveis consideram o recetor como rígido (maiores detalhes encontram-se descritos na referência ⁸⁵).

Relativamente à DM os softwares mais utilizados são: GROMACS, CHARMM, AMBER, NAMD e LAMMPS. Estes programas têm alguns recursos comuns e alguns recursos exclusivos, sendo alguns de código aberto (como por exemplo, GROMACS e LAMMPS).⁸⁶

No presente estudo, ambas as abordagens foram implementadas com o objetivo de explorar ao nível molecular o comportamento das NPS, selecionadas através da análise

quimiométrica levada a cabo na primeira fase do estudo, na presença do recetor alvo e de uma membrana biológica modelo. Especificamente, recorreu-se

- (i) ao *Docking* molecular, utilizando o pacote computacional Autodock 4.2, para estimar a forma preferencial com que as NPS interagem com o sítio ativo do recetor alvo;
- (ii) à simulação por dinâmica molecular, usando o programa GROMACS versão 2019, para estabelecer a correlação entre os perfis químicos identificados na análise quimiométrica referentes aos descritores moleculares e o efeito destas moléculas na estrutura de uma membrana lipídica modelo usada para mimetizar a barreira hematoencefálica (em inglês *blood-brain-barrier, BBB*).

2.4.1 Dinâmica Molecular

Ao contrário do *Docking* molecular, que tende a considerar conformações mais rígidas, do recetor e do ligando, a DM permite ultrapassar esta limitação ao conceder flexibilidade ao complexo proteína-ligando. Na base da DM está a aplicação da segunda Lei de Newton ao sistema molecular. O comportamento do sistema formado por um conjunto de átomos, por exemplo uma proteína, um complexo proteína-ligando, ou molécula inserida numa membrana, pode ser explorado durante um período específico de avaliação. Por outras palavras, a evolução temporal para um determinado conjunto de átomos que interagem entre si é seguida da integração da sua equação de movimento⁸², permitindo prever de que forma os átomos do sistema interagem ao longo do tempo, baseando-se num modelo físico que descreve essas interações.⁸⁶

Para tal, é necessário construir o sistema molecular, colocando dentro de uma caixa de simulação com um formato específico, na qual se incluem as moléculas cuja interação se pretende estudar e ainda moléculas com água, ou outro solvente caso se pretenda avaliar a interação em solução. Após passar por um processo de minimização de energia, o sistema é aquecido de forma gradual até uma temperatura desejada, promovendo o movimento dos átomos, ou seja, transformando a energia potencial em energia cinética. Ao longo do tempo as coordenadas espaciais e as velocidades dos átomos vão sofrendo alteração, até que a temperatura definida seja atingida. O sistema é então simulado considerando determinados parâmetros até atingir o equilíbrio e o seu comportamento é avaliado a partir da resolução da equação de movimento. Simultaneamente, a interação os átomos e as respetivas energias são

descritas a partir de um determinado campo de forças. A cada intervalo de tempo, e durante um determinado número de passos, a resultante do movimento e das forças que atuam no sistema é atualizada.

Na Figura 2.4 encontra-se uma representação esquemática simples de um procedimento de DM clássico.

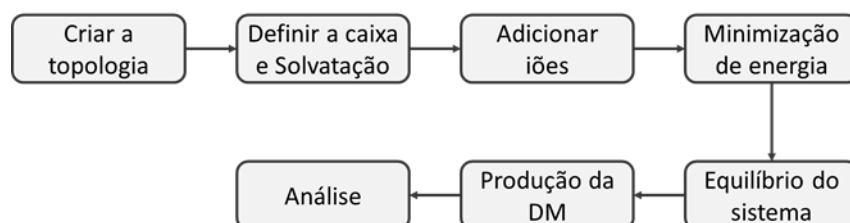


Figura 2.4 Esquema representativo das etapas principais da Dinâmica Molecular (DM) (adaptado de ⁸⁷).

Ensemble e Equações do Movimento

Numa simulação computacional é gerado um conjunto de configurações consecutivas de partículas, obtendo então um conjunto de microestados ordenados que estão sujeitos às mesmas condições termodinâmicas. A um conjunto de microestados dá-se o nome de *ensemble*. Consoante as restrições termodinâmicas a que o sistema está exposto, existem diferentes tipos de *ensemble*, entre eles:

- (i) *Ensemble* Canónico (NVT), onde N (número de partículas), V (volume) e T (temperatura) são mantidos constantes;
- (ii) *Ensemble* Microcanónico (NVE), no qual N, V e E (energia) são mantidas constantes;
- (iii) *Ensemble* Isotérmico-isobárico (NPT), onde mantidas constantes N, P (pressão) e T;
- (iv) Grande Canónico (μ VT), onde V, T e μ (potencial químico) são mantidos constantes.

Dado que as funções de estado são constantes, consoante o *ensemble*, é possível calcular as restantes propriedades do sistema através das relações fundamentais da termodinâmica.

As sucessivas configurações do sistema são obtidas pela integração da Lei de Movimento de Newton. O resultado consiste numa trajetória que especifica como é que as posições e velocidades das partículas do sistema variam com o tempo. A trajetória pode, então, ser obtida através da resolução da equação, na forma de equação diferencial, da segunda lei de Newton:

$$F_i = m_i a_i = m_i \frac{\partial^2 r_i}{\partial t^2} \quad (2.22)$$

onde m_i representa a massa da partícula, F_i a força, a_i é a aceleração e $\frac{\partial^2 r_i}{\partial t^2}$ é a segunda derivada da partícula na posição r . Esta equação descreve o movimento da partícula ao longo do eixo com força F_i .⁸⁸ Dado que não existe uma solução analítica para estas equações, este método recorre a Métodos Numéricos do tipo Diferenças Finitas. A ideia é dividir o passo da integração em pequenas etapas, separadas pelo tempo. A força total de cada partícula no tempo t é determinada através da soma vetorial das interações com outras partículas. Como vimos na equação 2.22, a partir da força podemos determinar a aceleração das partículas, que é depois combinada com a posição e velocidade no tempo t para o cálculo das mesmas para o tempo $t + \Delta t$. Isto acontece porque considera-se a força um termo constante.⁸⁸ Com isto e tendo em conta as coordenadas do átomo obtidas no intervalo $t_0 + \Delta t$ e a velocidade no instante $t_0 + \Delta t$, a posição e velocidade podem ser determinadas através das expressões:

$$\vec{r}_i(t_0 + \Delta t) = \vec{r}_i(t_0) + \vec{v}_i(t_0) \cdot \Delta t + \frac{1}{2} \vec{a}_i(t_0) \cdot \Delta t^2 \quad (2.23)$$

$$\vec{v}_i(t_0 + \Delta t) = \vec{v}_i(t_0) + \vec{a}_i(t_0) \cdot \Delta t \quad (2.24)$$

Dado que se considera a força constante, é essencial escolher o tempo de integração para a realização de uma simulação.

Escolha do tempo de integração

O tempo de integração numa simulação deve ser escolhido de acordo com o sistema que está a ser estudado. Um tempo de integração curto implica um aumento do tempo computacional necessário, diminuindo, conseqüentemente, o espaço conformacional a ser estudado nesse tempo. Por outro lado, um tempo de integração demasiado grande pode provocar erros na simulação devido a uma incorreta integração das equações de movimento. Para que não ocorram erros na simulação e os resultados obtidos sejam os mais exatos possível, torna-se essencial escolher o tempo de integração mais adequado para cada sistema. Na Tabela 2.2 encontra-se resumido o tempo de integração mais adequado para diferentes sistemas.⁸⁸

Tabela 2.2 Diferentes tipos de movimento presente em diversos sistemas e respetivos tempos de integração (adaptado de⁸⁸).

Sistema	Tipos de movimento	Tempo de integração (s)
Átomos	Translação	10^{-14}
Moléculas rígidas	Translação e rotação	5×10^{-15}

Moléculas flexíveis, ligações rígidas	Translação, rotação e torsão	2×10^{-15}
Moléculas flexíveis, ligações flexíveis	Translação, rotação, torsão e vibração	10^{-15} ou 5×10^{-16}

Pré-processamento da Dinâmica Molecular

Aquando da construção de um modelo de um sistema molecular, é necessário arranjar resposta a quatro aspetos fundamentais (Figura 2.5): (i) o número de graus de liberdade, atómicos ou moleculares, que deve, ser considerado explicitamente no sistema em estudo, (ii) o campo de forças a ser utilizado tendo em conta os graus de liberdade, (iii) como serão apresentados os graus de liberdade, (iv) como será efetuada a modelação das condições de fronteira e das forças externas.⁸⁹

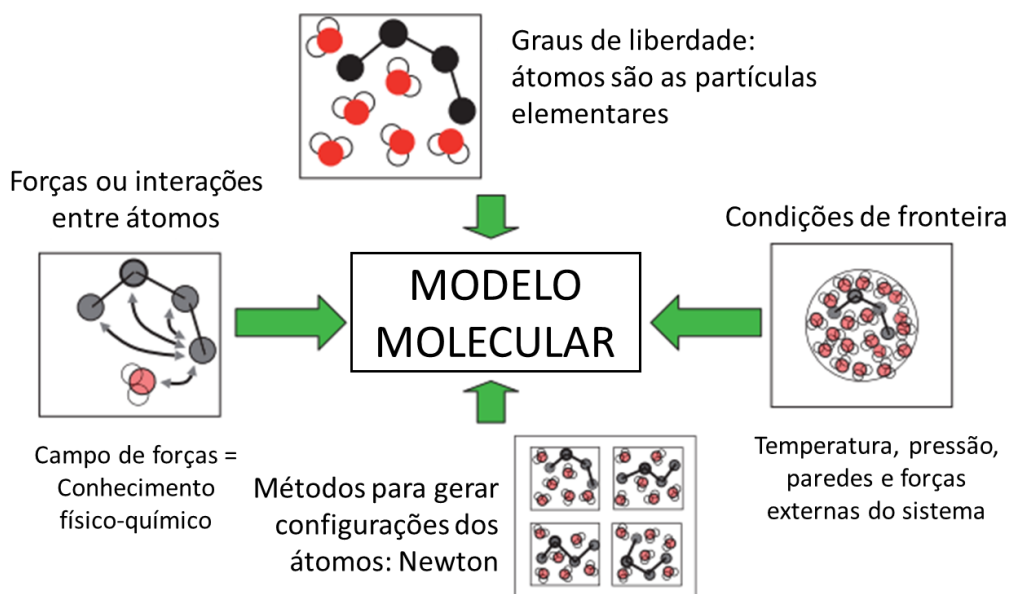


Figura 2.5 Diagrama representativo dos quatro tópicos necessários para desenvolver um modelo de simulação molecular (adaptado de ⁸⁹).

O campo de forças traduz-se por uma expressão matemática que engloba diferentes tipos de interação, nomeadamente as interações entre átomos ligados e não ligados.⁹⁰

No que diz respeito a interações entre átomos não ligados, a energia de interação pode ser descrita pelos potenciais de van der Waals e eletrostático. O potencial de van der Waals traduz a interação não eletrostática entre átomos não ligados. A repulsão entre átomos/moléculas a curtas distâncias é tida em conta usando potenciais de Lennard-Jones. Além destas há ainda que considerar a componente ou potencial eletrostático, entre átomos ou moléculas carregados que é tida em conta incluindo um potencial de Coulomb. A

contribuição das interações entre átomos diretamente ligados para a energia total tem em conta as componentes relativas aos modos de alongação, deformação angular e torção.

As simulações são realizadas numa caixa impondo condições fronteiras periódicas como forma de simular uma pequena porção do sistema. Assim, a caixa de simulação considera-se rodeada por cópias da mesma, anulando os efeitos de parede.

As condições de fronteira periódicas possibilitam que um número pequeno de partículas seja utilizado numa simulação. Este modelo cria réplicas em torno do sistema original e caso uma determinada partícula se desloque para o exterior da caixa de simulação, outra partícula igual entra no lado oposto. Na Figura 2.6 podemos observar um esquema representativo deste efeito.⁸⁸

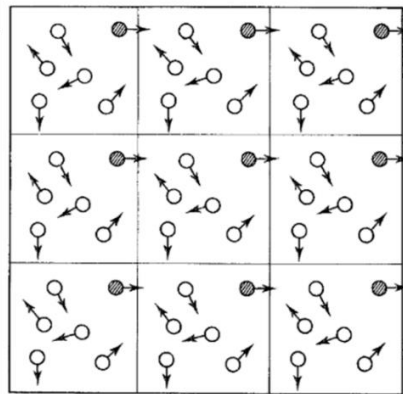


Figura 2.6 Esquema representativo do efeito da imposição de Condições de Fronteira Periódicas em duas dimensões (adaptado de ⁸⁸).

Análises nas simulações de DM

A fase final de uma simulação por DM é a obtenção e tratamento da informação relevante que permitirá fazer uma análise e caracterização do sistema. São várias as análises já implementadas no programa *Gromacs*, entre as quais se destacam a energia do sistema ao longo da simulação, rio, valores médios de temperatura, pressão e densidade, funções de distribuição radial, número de ligações de hidrogénio, raios de giração, desvio médio quadrático, entre outros. Em sistemas que incluem bicamadas lipídicas é possível determinar a densidade da membrana, a área ocupada por lípido, a espessura da bicamada, entre outros parâmetros. De seguida apresentam-se com maior detalhe as principais análises consideradas neste trabalho.

- Perfil de densidade⁹¹

Através do cálculo da densidade de cada átomo ou molécula ao longo da bicamada é possível determinar o posicionamento relativo quer das componentes da membrana quer o

posicionamento preferencial dos compostos no interior da bicamada. Habitualmente são selecionados determinados grupos de átomos da bicamada que representem zonas mais exteriores como por exemplo os átomos P e N dos grupos polares localizados nas cabeças dos lípidos (região hidrofílica) e ainda os grupos metilo terminais localizados nas extremidades das cadeias hidrofóbicas que representam zonas mais internas da bicamada. Neste trabalho as NPS foram representadas pelo seu centro de massa.

- Parâmetro de Ordem (S_{CD})⁹²

O parâmetro de ordem (S_{CD}) permite avaliar o grau de ordenação das cadeias lipídicas da bicamada e o seu valor é obtido através da expressão:

$$S_{CD} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle \quad (2.25)$$

onde θ é o ângulo entre o vetor C-H selecionado e a direção de referência (bicamada). Quanto mais próximo o valor de S_{CD} estiver de 0 (zero), menor é o alinhamento das cadeias alifáticas. Numa simulação onde se utiliza um sistema de átomos unidos, é possível determinar S_{CD} a partir das posições da cadeia alifática, assumindo que os ângulos de ligação dos hidrogénio estão num estado de equilíbrio. Assim, o parâmetro de ordem, tensor (S), pode ser definido como:

$$S_{ij} = \frac{1}{2} \langle 3 \cos \theta_i \theta_j - \delta_{ij} \rangle \quad (2.26)$$

onde θ_i é o ângulo relativo entre o eixo molecular i e a normal à bicamada. Os parênteses representam a média sobre as moléculas e o tempo de simulação e δ_{ij} é o delta de Kronecker.

Em qualquer carbono não terminal da cadeia alifática da membrana, C_n , o sistema do eixo molecular pode ser definido usando átomos de carbono vizinhos da cadeia: considera-se o eixo z como o vetor normalizado $\overrightarrow{C_{n-1}C_{n+1}}$, o eixo x é o vetor unitário perpendicular a $\overrightarrow{C_{n-1}C_n}$ e $\overrightarrow{C_nC_{n+1}}$ e o eixo y é o produto vetorial dos eixos z e x . Nas simulações deste estudo é utilizado um campo de forças de átomos unidos, para tal, o parâmetro de ordem é determinado através das relações:

$$S_{CD}^{sat} = \frac{3}{2} S_{xx} + \frac{1}{3} S_{yy} \quad (2.27)$$

$$S_{CD}^{Insat} = \frac{1}{4} S_{zz} + \frac{3}{4} S_{yy} + \frac{\sqrt{3}}{2} S_{xx} \quad (2.28)$$

Através da comparação dos valores deste parâmetro entre o sistema em estudo e um sistema de referência, como é o caso de uma bicamada sem NSP, é possível avaliar se a presença da droga promove ou não uma desordenação da bicamada.

- Difusão lateral da bicamada⁹¹

A difusão lateral, determinada através do deslocamento quadrático médio (MSD, em inglês *Mean Square Displacement*), permite medir o desvio da membrana em relação a uma posição de referência ao longo do tempo. Numa primeira instância, mede-se o MSD das moléculas individuais a partir das trajetórias da DM e, depois, calcula-se pela fórmula de Einstein em duas dimensões através da expressão:

$$D = \frac{1}{2d} \lim_{t \rightarrow \infty} \frac{dMSD(t)}{dt} \quad (2.29)$$

Sendo que, a determinação de MSD(t) é definida como:

$$MSD(t) = \langle \|\vec{R}(t + t_0) - \vec{R}(t_0)\|^2 \rangle \quad (2.30)$$

onde \vec{R} é o vetor posição do centro de massa da bicamada e a média é calculada relativamente ao tempo com origem em t_0 . Para eliminar o ruído provocado por flutuações, considerou-se o centro de massa das membranas fixo aquando das análise de MSD nos diferentes sistemas (bicamada e bicamada com NSP).

2.4.2 Docking molecular

O *Docking* é um método de simulação computacional que permite prever o modo como um ligando interage com um dado alvo, recetor, proteína, DNA, etc., com o objetivo de estabelecer o modo de ligação mais favorável e a sua relação com a atividade biológica do ligando. A energia de ligação estimada ou as funções de pontuação traduzem a afinidade dessa interação. A interação, geralmente, envolve uma proteína e uma molécula de pequenas dimensões. Contudo, também é possível utilizar este método para avaliar a interação entre duas proteínas.⁹³

Dois conceitos ligados a este tipo de simulação são: (i) *Dock pose*, este termo é utilizado para referir uma posição, conformação ou orientação entre um recetor e um ligando; (ii) *Dock score* é um valor criado a partir de um conjunto de parâmetros que permite avaliar a compatibilidade entre o recetor e o ligando. Um bom *Dock score* irá corresponder à afinidade favorável entre o ligando e o recetor.⁹⁴ A Figura 2.7 descreve esquematicamente estes conceitos.

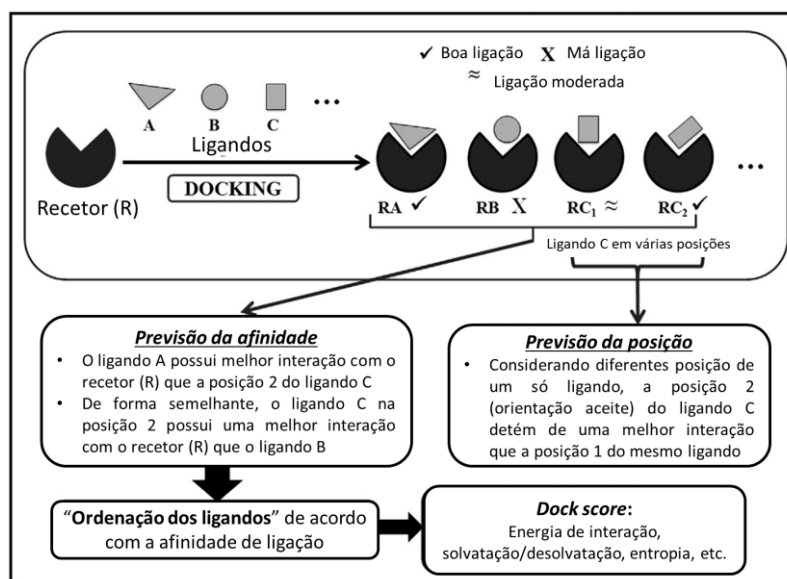


Figura 2.7 Esquema-resumo dos pontos fundamentais no estudo do *Docking* molecular (adaptado de ⁹⁴).

Resumidamente, podemos afirmar que o principal objetivo do *Docking* molecular é estabelecer qual a melhor orientação estrutural do ligando relativamente ao recetor e a respetiva energia de ligação. Além disso, perante um conjunto de ligandos, é possível compará-los utilizando os mesmos parâmetros.

Desafios do *Docking*

Um dos desafios deste método é a presença de moléculas de água na estrutura disponível do recetor. Tendo em conta que o apresenta muitas vezes um papel importante na interação entre um recetor e um ligando, se a sua presença for ignorada, a energia de ligação pode ser muito baixa. Por outro lado, se a simulação for realizada na presença do solvente, em particular no caso da água, as moléculas de água poderão ligar-se ao recetor por ligações de hidrogénio fortes, impedindo a interação ligando-recetor.⁹⁵

Os recetores flexíveis podem ser desafiantes para o *Docking* molecular, pois a possibilidade de tomarem diferentes conformações pode dificultar a simulação com o ligando, aumentar o espaço de pesquisa, o tempo e os recursos computacional necessários. Uma das alternativas para ultrapassar este problema, é a utilização de um programa que considere o recetor como uma estrutura rígida. É necessário ter, também, a estrutura 3D do recetor determinada experimentalmente, o que nem sempre acontece constituindo, por isso, muitas vezes uma limitação a este tipo de estudos.

Etapas na aplicação do *Docking* molecular

O processo do *Docking* molecular envolve um conjunto de etapas que serão enumeradas a seguir. A Figura 2.8 ilustra, resumidamente, os diferentes passos utilizados neste método.

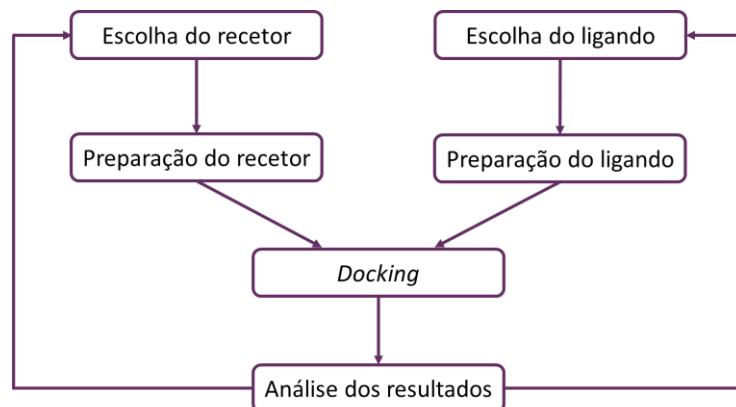


Figura 2.8 Diagrama representativo do conjunto de etapas no *Docking* molecular (adaptado de ⁹³).

Obtenção das coordenadas 3D do recetor

As coordenadas espaciais do recetor podem ser obtidas através da plataforma *The Protein Data Bank (PDB)* através do site <https://www.rcsb.org/>, onde é disponibilizada a informação sobre as coordenadas x, y, z dos átomos, o tipo de molécula (proteína, solvente ou ligando), tipo de resíduo, dados sobre as condições experimentais, a estrutura secundária e quaternária, entre outros.

Após a obtenção das coordenadas do recetor, é necessário prepará-lo para a simulação. Este passo consiste em retirar dados que não são necessários para a simulação, como por exemplo as moléculas de água, eventuais ligandos, cofatores ou outras moléculas presentes.

Uma vez obtidas as coordenadas do ligando e do recetor o passo seguinte é o de definir a caixa onde irá decorrer a simulação. A definição da caixa de simulação, ou seja, a janela espacial onde irá ser explorada a interação ligando-recetor pode ser feita considerando o centroide do ligando ou o local ativo do recetor, caso seja conhecido.

Uma vez terminada a simulação, é possível extrair um conjunto de informações que nos permitirão fazer uma análise ao sistema, nomeadamente a energia de ligação e representações gráfica das interações que existem em cada *Dock pose*.

2.4.3 Método do Gradiente Independente

O Método do Gradiente Independente assenta na análise da densidade da carga eletrónica e respetivos gradientes, permitindo visualizar e quantificar as regiões de interação

estabilizantes/destabilizantes entre duas moléculas. Esta análise rege-se por características topológicas da densidade de carga eletrónica (ρ), de cada sistema, e utiliza quantidades correspondentes à primeira e segunda derivadas da densidade.^{96,97}

O termo δg^{inter} diz respeito ao descritor de IGM dado à diferença entre as primeiras derivada das densidades de carga para o sistema total e os fragmentos,⁹⁸

$$\delta g^{inter} = |\nabla \rho^{IGM,inter}| - |\nabla \rho| \quad (2.31)$$

Quando o valor desta diferença é positivo, significa que estão presentes interações não covalentes. A magnitude do descritor num ponto do espaço permite deduzir a força da interação. $\nabla \rho^{IGM,inter}$ é obtida através do somatório de todos os N átomos dos diferentes fragmentos. A expressão 2.32 demonstra a sua expressão, neste caso para o eixo x e átomos indicados pelas letras A e B.⁹⁸

$$\left(\frac{\delta \rho}{\delta x}\right)^{IGM,inter} = \left| \sum_{i=1}^{N_A} \frac{\delta \rho_i}{\delta x} \right| + \left| \sum_{i=1}^{N_B} \frac{\delta \rho_i}{\delta x} \right| \quad (2.32)$$

O IGMPlot, ferramenta utilizada para visualização das interações não covalentes, utiliza densidades de cargas atómicas pré-calculadas para estimar a densidade molecular que produz o efeito mínimo nas interações não covalentes, uma vez que ocorre um relaxamento mínimo das densidades de carga nas regiões de baixa densidade, seguida da formação de ligações covalentes. Apesar de ser possível identificar regiões com interações não covalentes utilizando os valores de δg^{inter} , é necessário a segunda derivada (*Laplacian*) da densidade $\nabla^2 \rho$ para discriminar as interações favoráveis e não favoráveis. É possível obter informação acerca das interações estabilizantes ($\lambda_2 < 0$) e desestabilizantes ($\lambda_2 > 0$) através da decomposição do termo de *Laplacian* em $\nabla^2 \rho = \lambda_1 + \lambda_2 + \lambda_3$ ($\lambda_1 \leq \lambda_2 \leq \lambda_3$), os três valores próprios da variação máxima. Interações mais fortes, como é o caso das ligações de hidrogénio, apresentam valores maiores de $signal(\lambda_2)\rho$. Interações mais fracas, como é o caso das forças van der Waals apresentam valores próximos de 0. Numa representação gráfica, as forças estabilizadoras surgem no lado negativo do eixo do x e as formas desestabilizadoras no lado positivo.⁹⁹

Esta análise foi usada neste trabalho para identificar o tipo e a intensidade de interações não covalentes existentes entre as NSP selecionadas e um recetor apropriado.

Capítulo 3

Descrição da base de dados

3. Descrição da base de dados

Neste capítulo será descrito o conjunto de dados que foi alvo deste estudo, bem como os programas e ferramentas de análise utilizadas na sua racionalização. A primeira secção deste capítulo centra-se na descrição do procedimentos de aquisição e pré-processamento da informação estrutural referente às NSP selecionadas para o estudo. Na secção seguinte, encontra-se detalhadas os descritores moleculares usados na sua discriminação, bem como os programas e os parâmetros necessários à caracterização da estrutura e dinâmica das NSP.

3.1 Coordenadas 3D das NPS

O conjunto de dados construído para a análise estrutural das NSP contém um total de 301 moléculas organizadas em dois grandes grupos: Controladas (C) e Não Ilícitas (NI). No que diz respeito ao grupo C, as moléculas que fazem parte deste grupo já foram estudadas, são controladas pelo UNODC, (em inglês *United Nations Office on Drugs and Crime*) e muitas delas constam na legislação. Já o grupo NI refere-se a substâncias psicoativas que foram pouco estudadas, não são controladas e o seu uso não está ainda legislado ou regulamentado. As moléculas consideradas no conjunto de dados pertencem a 9 famílias: Aminoindanos, Canabinoides Sintéticos, Catinonas Sintéticas, Fenetilaminas, Novas Benzodiazepinas, Opióides Sintéticos, Piperazinas, Substâncias tipo Fenciclidinas e Triptaminas. Adicionalmente, as moléculas que não estavam alocadas a nenhuma destas famílias foram categorizadas como Outras drogas.

As moléculas pertencentes ao grupo C constam no documento redigido pelo INCB (em inglês *International Narcotics Control Board*) que tem como título *List of Psychotropic Substances under International Control*¹⁰⁰ apresentado em 2020 e/ou estão presentes na Portaria n.º 154/2013, de 17 de Abril ao abrigo da legislação portuguesa. Por outro lado, as moléculas pertencentes ao grupo NI não constam nos documentos mencionados acima, mas sim em artigos científicos^{39,42,44,46,48,57}, relatórios emitidos pela EMCDDA, como é o caso da NPS MDMB-4en-PINACA¹⁰¹, que pertence à família dos Canabinoides Sintéticos, e/ou nas publicações lançadas na página oficial da UNODC na secção *Latest News*¹⁰².

Para a realização deste estudo, a estrutura das 301 moléculas foi construída e otimizada ao nível da Mecânica Molecular no *software* Avogadro (versão 1.2.0).

De forma a simplificar a descrição e sistematização da informação sobre as famílias das drogas consideradas, foi adotado ao longo do estudo, um esquema de cores para representar as diferentes famílias. Nas Tabelas 3.1 a 3.10 encontram-se listadas as drogas pertencentes a cada família bem como a cor correspondente.

Neste trabalho, considerou-se como drogas de controlo aquelas que já foram classificadas como ilícitas, as drogas “Não ilícitas” correspondem às NPS sobre as quais não existe qualquer legislação.

Tabela 3.1 Lista de drogas pertencentes à família dos Aminoindanes divididos pelas categorias Controlo (C) e Não ilícitas (NI).

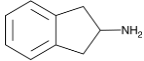
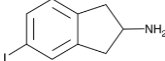
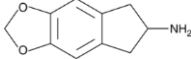
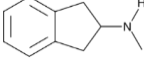
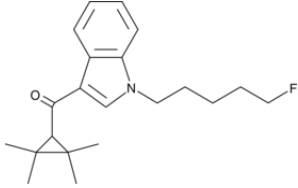
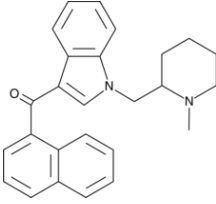
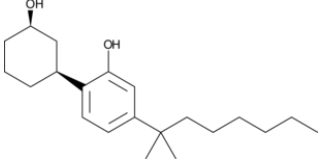
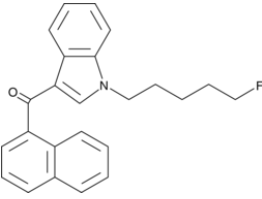
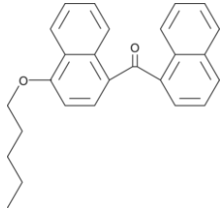
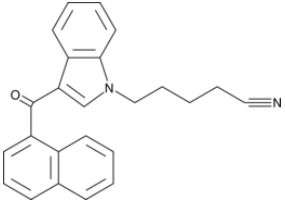
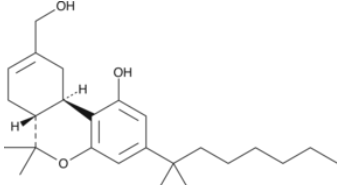
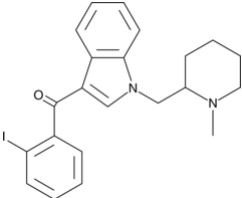
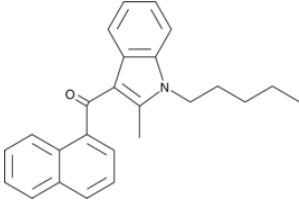
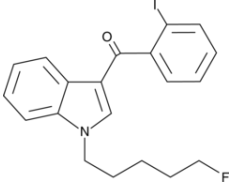
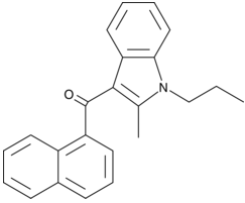
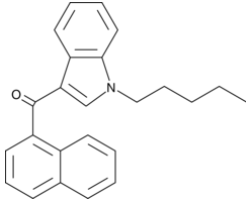
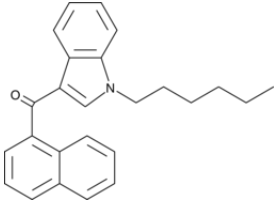
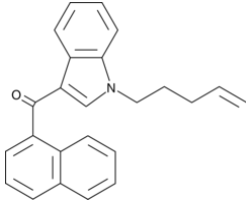
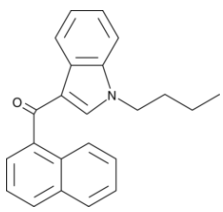
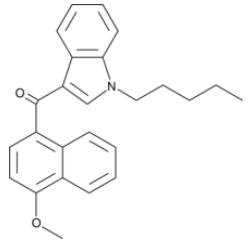
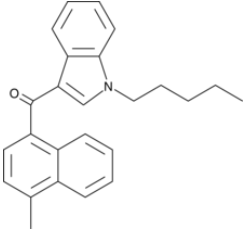
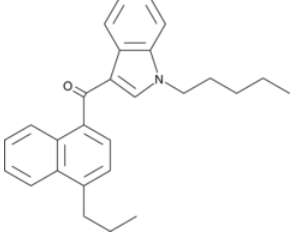
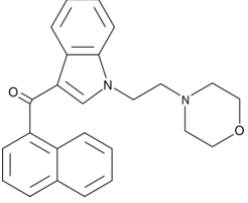
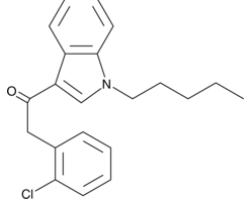
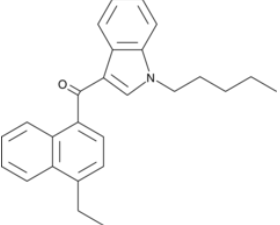
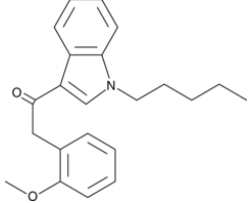
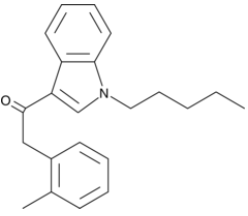
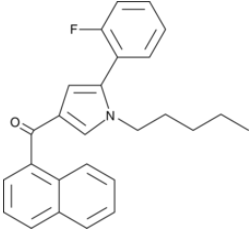
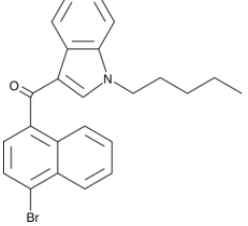
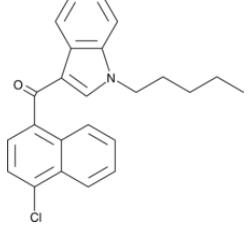
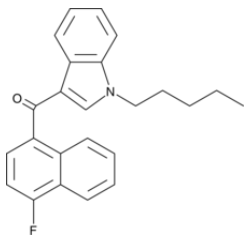
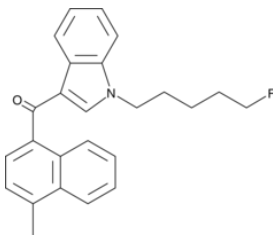
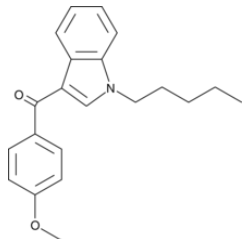
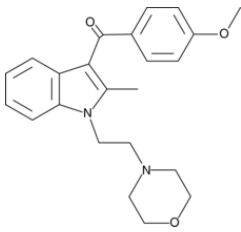
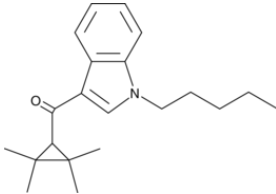
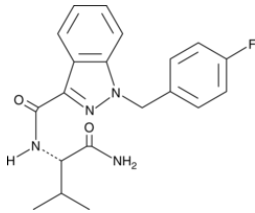
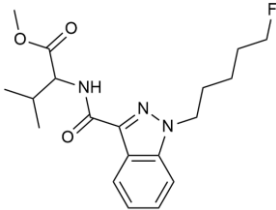
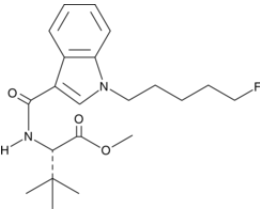
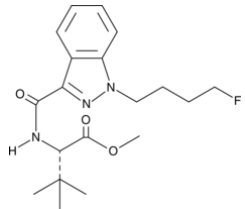
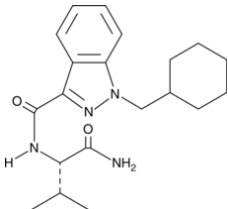
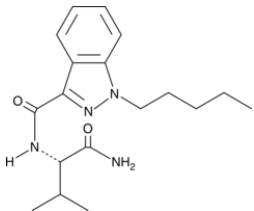
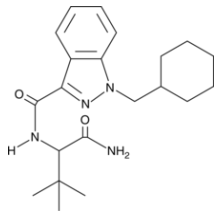
Aminoindanes			
Controlo	Abreviatura	Controlo	Abreviatura
2-AI 	A1_C	5-IAI 	A2_C
MDAI 	A3_C		
Não ilícito	Abreviatura		
NM-2AI 	A4_NI		

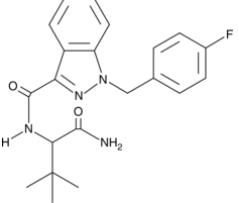
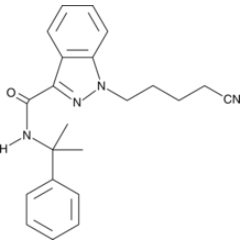
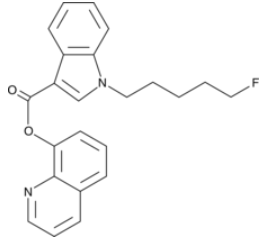
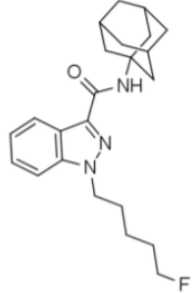
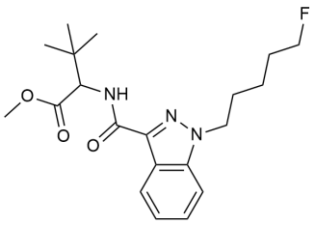
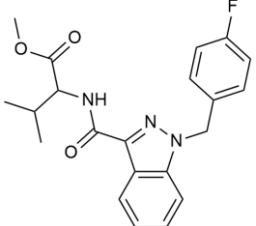
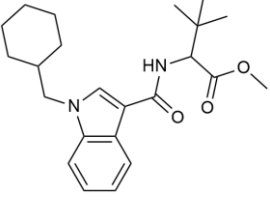
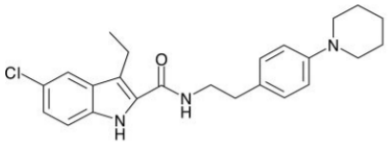
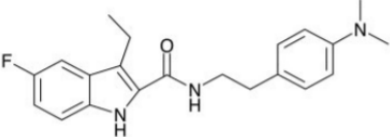
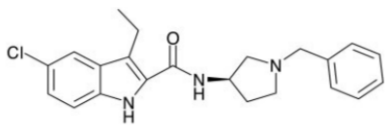
Tabela 3.2 Lista de drogas pertencentes à família dos Canabinoides Sintéticos divididos pelas categorias Controlo (C) e Não ilícitas (NI).

Canabinoides Sintéticos			
Controlo	Abreviatura	Controlo	Abreviatura
5-FUR-144 	CS1_C	AM-1220 	CS2_C

<p>CP 47,497</p> 	<p>CS3_C</p>	<p>AM-2201</p> 	<p>CS4_C</p>
<p>CRA-13</p> 	<p>CS5_C</p>	<p>AM-2232</p> 	<p>CS6_C</p>
<p>HU-210</p> 	<p>CS7_C</p>	<p>AM-2233</p> 	<p>CS8_C</p>
<p>JWH-007</p> 	<p>CS9_C</p>	<p>AM-694</p> 	<p>CS10_C</p>
<p>JWH-015</p> 	<p>CS11_C</p>	<p>JWH-018</p> 	<p>CS12_C</p>
<p>JWH-019</p> 	<p>CS13_C</p>	<p>JWH-022</p> 	<p>CS14_C</p>

<p style="text-align: center;">JWH-073</p> 	CS15_C	<p style="text-align: center;">JWH-081</p> 	CS16_C
<p style="text-align: center;">JWH-122</p> 	CS17_C	<p style="text-align: center;">JWH-182</p> 	CS18_C
<p style="text-align: center;">JWH-200</p> 	CS19_C	<p style="text-align: center;">JWH-203</p> 	CS20_C
<p style="text-align: center;">JWH-210</p> 	CS21_C	<p style="text-align: center;">JWH-250</p> 	CS22_C
<p style="text-align: center;">JWH-251</p> 	CS23_C	<p style="text-align: center;">JWH-307</p> 	CS24_C
<p style="text-align: center;">JWH-387</p> 	CS25_C	<p style="text-align: center;">JWH-398</p> 	CS26_C

<p>JWH-412</p> 	<p>CS27_C</p>	<p>MAM-2201</p> 	<p>CS28_C</p>
<p>RCS-4</p> 	<p>CS29_C</p>	<p>WIN-48,098</p> 	<p>CS30_C</p>
<p>UR-144</p> 	<p>CS31_C</p>	<p>AB-FUBINACA</p> 	<p>CS32_C</p>
<p>5F-AMB-PINACA</p> 	<p>CS33_C</p>	<p>5F-MDMB-PICA</p> 	<p>CS34_C</p>
<p>4F-MDMB-BINACA</p> 	<p>CS35_C</p>	<p>AB-CHMINACA</p> 	<p>CS36_C</p>
<p>AB-PINACA</p> 	<p>CS37_C</p>	<p>ADB-CHMINACA</p> 	<p>CS38_C</p>

<p>ADB-FUBINACA</p> 	<p>CS39_C</p>	<p>CUMYL-4CN-BINACA</p> 	<p>CS40_C</p>
<p>5F-PB-22</p> 	<p>CS41_C</p>	<p>AF-APINACA</p> 	<p>CS42_C</p>
<p>5F-MDMB-PINACA</p> 	<p>CS43_C</p>	<p>FUB-AMB</p> 	<p>CS44_C</p>
<p>MDMB-CHMICA</p> 	<p>CS45_C</p>	<p>Org 27569</p> 	<p>CS46_C</p>
<p>Org 27759</p> 	<p>CS47_C</p>	<p>Org 29647</p> 	<p>CS48_C</p>

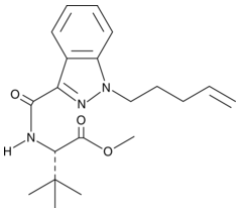
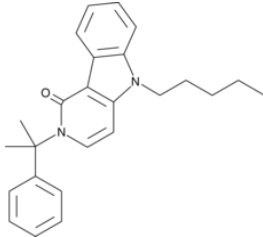
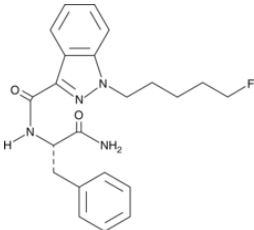
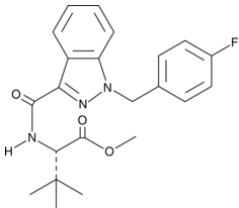
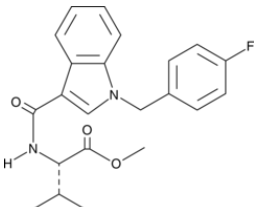
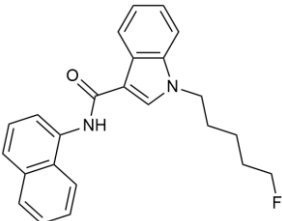
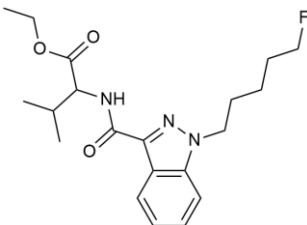
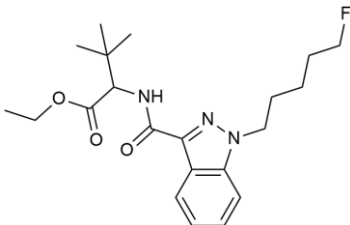
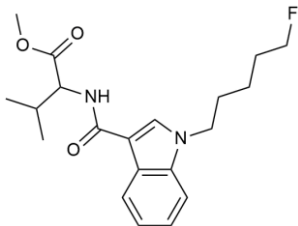
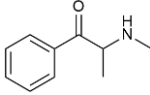
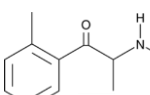
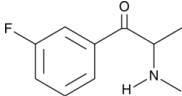
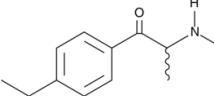
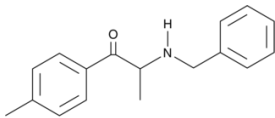
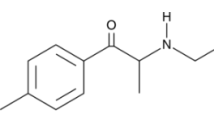
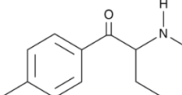
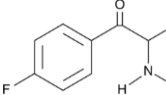
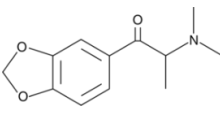
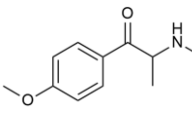
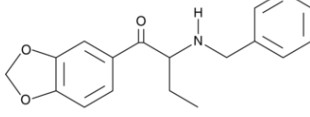
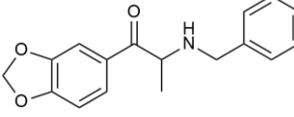
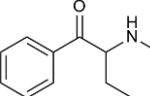
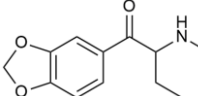
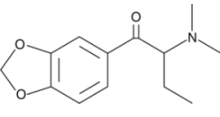
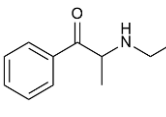
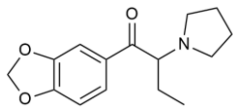
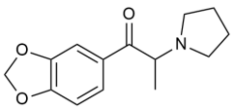
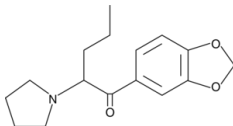
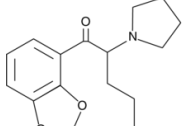
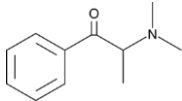
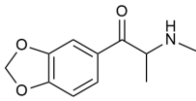
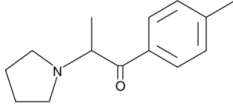
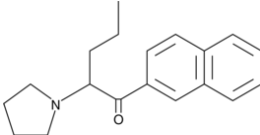
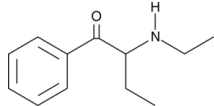
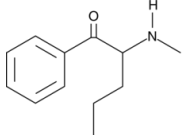
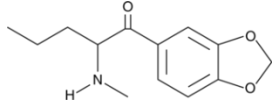
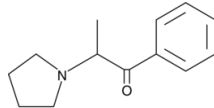
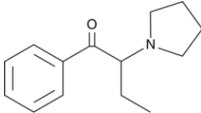
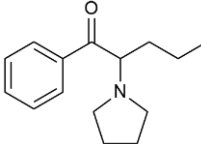
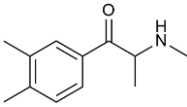
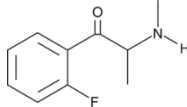
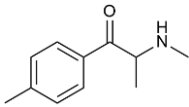
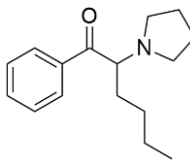
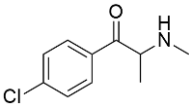
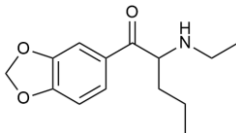
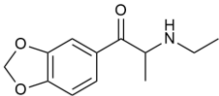
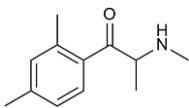
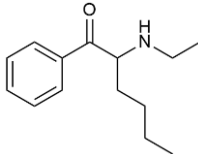
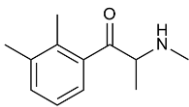
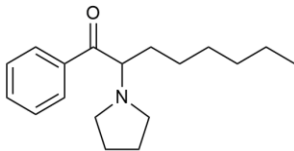
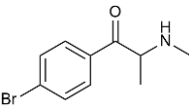
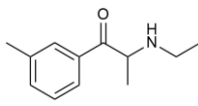
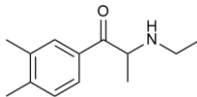
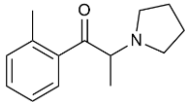
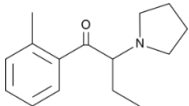
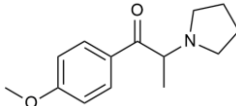
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p data-bbox="280 282 544 315">MDMB-4en-PINACA</p> 	<p data-bbox="671 394 782 427">CS49_NI</p>	<p data-bbox="903 271 1171 304">CUMYL-PeGACLONE</p> 	<p data-bbox="1291 394 1401 427">CS50_NI</p>
<p data-bbox="379 600 448 633">PX-2</p> 	<p data-bbox="671 719 782 752">CS51_NI</p>	<p data-bbox="916 611 1158 645">MDMB-FUBINACA</p> 	<p data-bbox="1291 719 1401 752">CS52_NI</p>
<p data-bbox="325 920 501 954">AMB-FUBICA</p> 	<p data-bbox="671 1028 782 1061">CS53_NI</p>	<p data-bbox="967 909 1107 943">5F-MN-24</p> 	<p data-bbox="1291 1028 1401 1061">CS54_NI</p>
<p data-bbox="304 1216 521 1249">5F-EBM-PINACA</p> 	<p data-bbox="671 1335 782 1368">CS55_NI</p>	<p data-bbox="919 1216 1155 1249">5F-EDMB-PINACA</p> 	<p data-bbox="1291 1335 1401 1368">CS56_NI</p>
<p data-bbox="320 1523 505 1556">5F-MMB-PICA</p> 	<p data-bbox="671 1641 782 1675">CS57_NI</p>		

Tabela 3.3 Lista de drogas pertencentes à família das Catinonas Sintéticas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

Catinonas Sintéticas			
Controlo	Abreviatura	Controlo	Abreviatura
Metcatinona 	KS1_C	2-Metilmetcatinona 	KS2_C
3-FMC 	KS3_C	4-EMC 	KS4_C
4-MBC 	KS5_C	4-MEC 	KS6_c
4-methylbuphedrone 	KS7_C	4-FMC 	KS8_C
bk-MDDMA 	KS9_C	Metedrona 	KS10_C
BMDDB 	KS11_C	BMDP 	KS12_C
Bufedrona 	KS13_C	bk-MBDB 	KS14_C
bk-MMBDB 	KS15_C	Etcatinona 	KS16_C

<p>3,4-MDPBP</p> 	KS17_C	<p>3,4-MDPPP</p> 	KS18_C
<p>3,4-MDPV</p> 	KS19_C	<p>2,3-MDPV</p> 	KS20_C
<p>N,N-DMC</p> 	KS21_C	<p>Metilona</p> 	KS22_C
<p>4-metil PPP</p> 	KS23_C	<p>Nafirone</p> 	KS24_C
<p>NEB</p> 	KS25_C	<p>Pentedrona</p> 	KS26_C
<p>Pentilon</p> 	KS27_C	<p>α-PPP</p> 	KS28_C
<p>α-PBP</p> 	KS29_C	<p>α-PVP</p> 	KS30_C
<p>3,4-DMMC</p> 	KS31_C	<p>2-FMC</p> 	KS32_C

Mefedrona 	KS33_C	α-PHP 	KS34_C
4-CMC 	KS35_C	N-Etilnorpentilona 	KS36_C
Etilona 	KS37_C		
Não ilícito	Abreviatura	Não ilícito	Abreviatura
2,4-DMMC 	KS38_NI	NEH 	KS39_NI
2,3-DMMC 	KS40_NI	PV9 	KS41_NI
4-BMC 	KS42_NI	3-MEC 	KS43_NI
2,3-DMEC 	KS44_NI	2-MePPP 	KS45_NI
2-MePBP 	KS46_NI	MOPPP 	KS47_NI

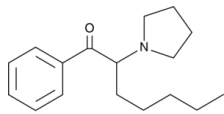
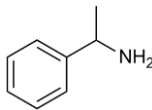
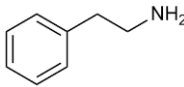
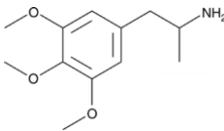
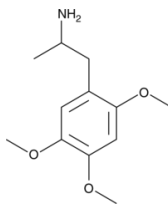
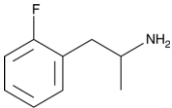
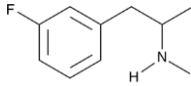
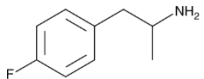
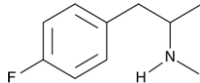
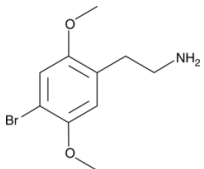
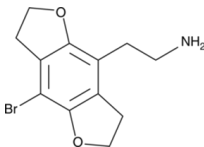
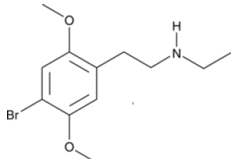
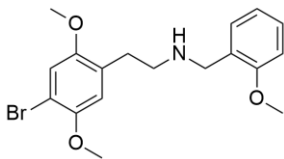
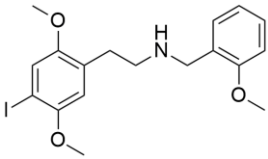
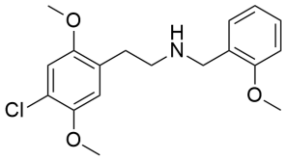
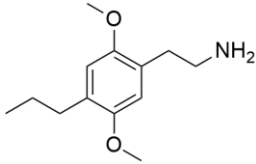
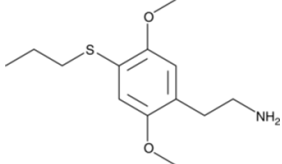
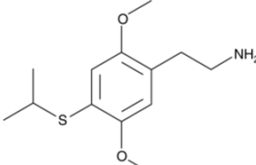
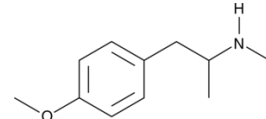
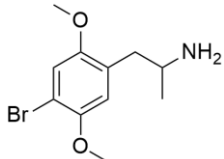
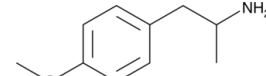
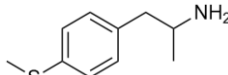
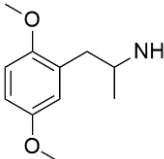
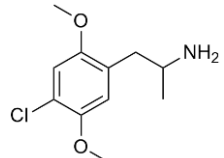
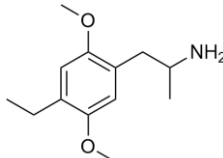
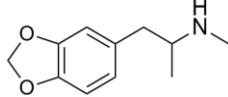
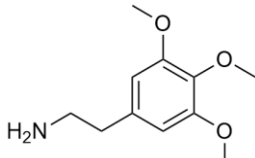
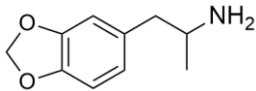
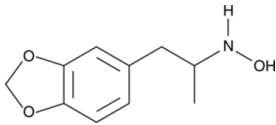
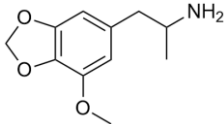
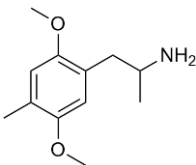
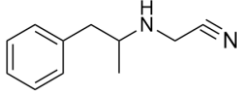
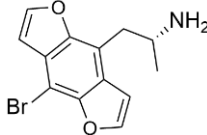
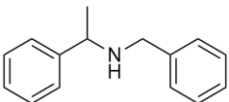
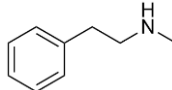
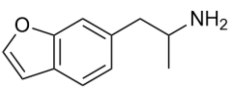
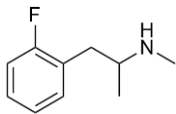
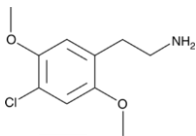
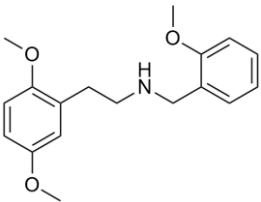
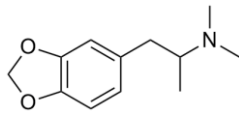
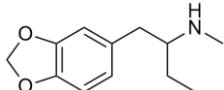
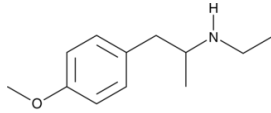
<p>PV8</p> 	<p>KS48_NI</p>		
--	----------------	--	--

Tabela 3.4 Lista de drogas pertencentes à família das Fenetilaminas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

Fenetilaminas			
Controlo	Abreviatura	Controlo	Abreviatura
<p>1-PEA</p> 	<p>Ph1_C</p>	<p>2-PEA</p> 	<p>Ph2_C</p>
<p>TMA</p> 	<p>Ph3_C</p>	<p>TMA-2</p> 	<p>Ph4_C</p>
<p>2-FA</p> 	<p>Ph5_C</p>	<p>3-FMA</p> 	<p>Ph6_C</p>
<p>4-FA</p> 	<p>Ph7_C</p>	<p>4-FMA</p> 	<p>Ph8_C</p>
<p>2C-B</p> 	<p>Ph9_C</p>	<p>2C-B-FLY</p> 	<p>Ph10_C</p>
<p>N-Etil 2C-B</p> 	<p>Ph11_C</p>	<p>25B-NBOMe</p> 	<p>Ph12_C</p>

<p>25I-NBOMe</p> 	Ph13_C	<p>25C-NBOMe</p> 	Ph14_C
<p>2C-P</p> 	Ph15_C	<p>2C-T-7</p> 	Ph16_C
<p>2C-T-4</p> 	Ph17_C	<p>PMMA</p> 	Ph18_C
<p>DOB</p> 	Ph19_C	<p>PMA / 4-MA</p> 	Ph20_C
<p>4-MTA</p> 	Ph21_C	<p>DMA</p> 	Ph22_C
<p>DOC</p> 	Ph23_C	<p>DOET</p> 	Ph24_C
<p>MDMA</p> 	Ph25_C	<p>Mescaline</p> 	Ph26_C

<p>MDA</p> 	Ph27_C	<p>N-Hidroxi MDA</p> 	Ph28_C
<p>MMDA</p> 	Ph29_C	<p>DOM</p> 	Ph30_C
<p>Fenproporex</p> 	Ph31_C	<p>Bromo-Dragonfly</p> 	Ph32_C
<p>N-benzil 1-fenetilamina</p> 	Ph33_C	<p>NMPEA</p> 	Ph34_C
<p>6-APB</p> 	Ph35_C		
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>2-FMA</p> 	Ph36_NI	<p>2C-C</p> 	Ph37_NI
<p>25H-NBOMe</p> 	Ph38_NI	<p>MDDM</p> 	Ph39_NI
<p>MBDB</p> 	Ph40_NI	<p>PMEA</p> 	Ph41_NI

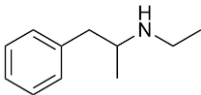
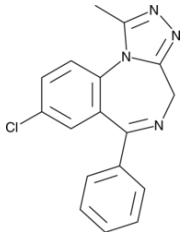
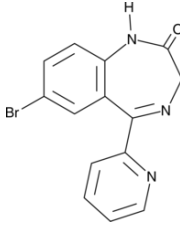
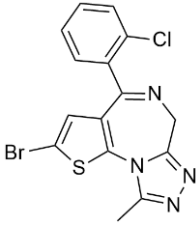
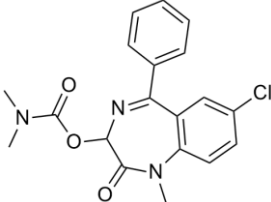
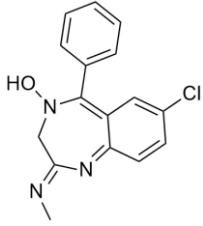
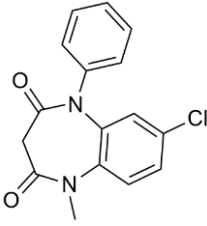
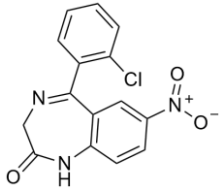
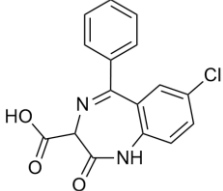
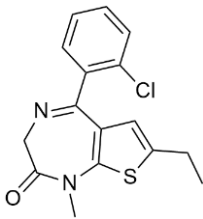
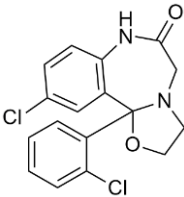
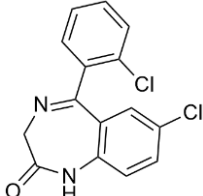
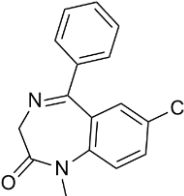
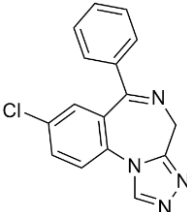
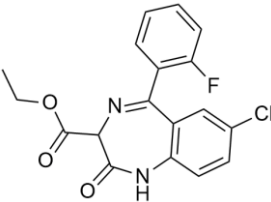
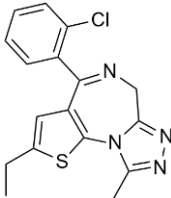
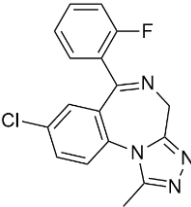
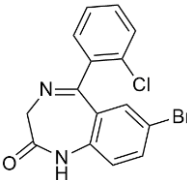
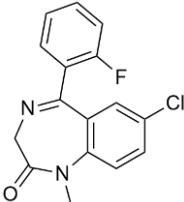
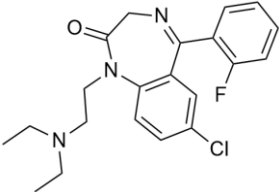
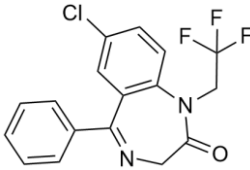
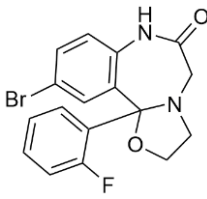
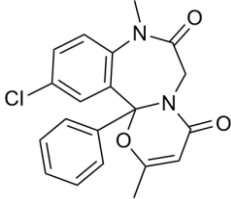
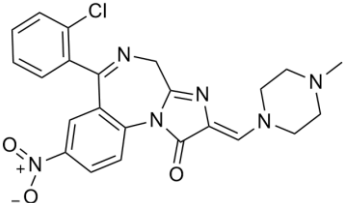
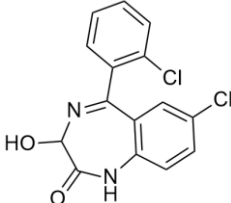
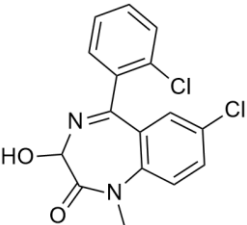
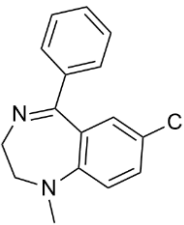
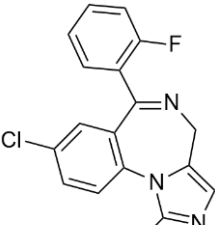
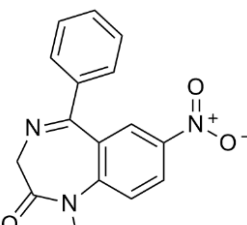
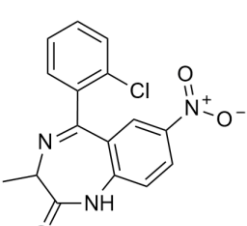
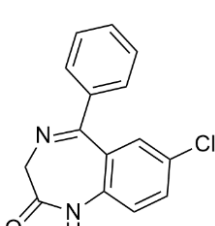
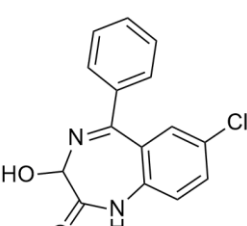
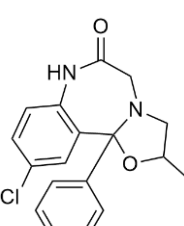
<p>N-etilamfetamina</p> 	<p>Ph42_NI</p>		
---	----------------	--	--

Tabela 3.5 Lista de drogas pertencentes à família das Novas Benzodiazepinas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

<p style="text-align: center;">Novas Benzodiazepinas</p>			
<p>Controlo</p>	<p>Abreviatura</p>	<p>Controlo</p>	<p>Abreviatura</p>
<p>Alprazolam</p> 	<p>NB1_C</p>	<p>Bromazepam</p> 	<p>NB2_C</p>
<p>Brotizolam</p> 	<p>NB3_C</p>	<p>Camazepam</p> 	<p>NB4_C</p>
<p>Clordiazepoxide</p> 	<p>NB5_C</p>	<p>Clobazam</p> 	<p>NB6_C</p>
<p>Clonazepam</p> 	<p>NB7_C</p>	<p>Clorazepate</p> 	<p>NB8_C</p>

<p>Clotiazepam</p> 	NB9_C	<p>Cloxazolam</p> 	NB10_C
<p>Delorazepam</p> 	NB11_C	<p>Diazepam</p> 	NB12_C
<p>Estazolam</p> 	NB13_C	<p>Etil loflazepate</p> 	NB14_C
<p>Etizolam</p> 	NB15_C	<p>Flualprazolam</p> 	NB16_C
<p>Fenazepam</p> 	NB17_C	<p>Fludiazepam</p> 	NB18_C
<p>Flurazepam</p> 	NB19_C	<p>Halazepam</p> 	NB20_C

<p>Haloxazolam</p> 	NB21_C	<p>Ketazolam</p> 	NB22_C
<p>Loprazolam</p> 	NB23_C	<p>Lorazepam</p> 	NB24_C
<p>Lormetazepam</p> 	NB25_C	<p>Medazepam</p> 	NB26_C
<p>Midazolam</p> 	NB27_C	<p>Nimetazepam</p> 	NB28_C
<p>Meclonazepam</p> 	NB29_C	<p>Nordazepam</p> 	NB30_C
<p>Oxazepam</p> 	NB31_C	<p>Oxazolam</p> 	NB32_C

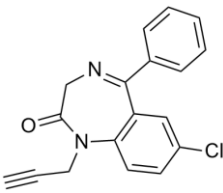
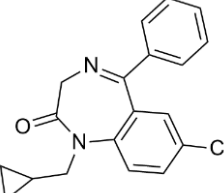
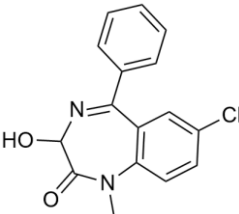
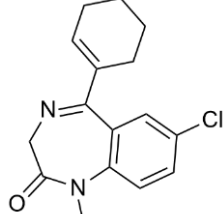
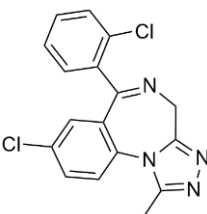
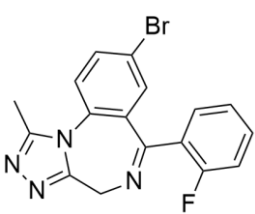
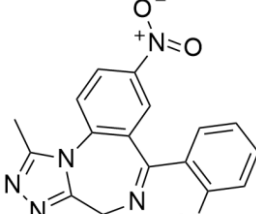
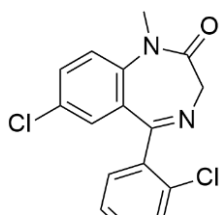
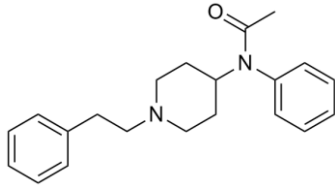
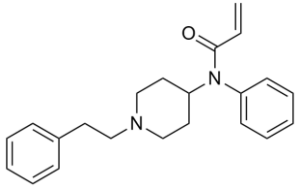
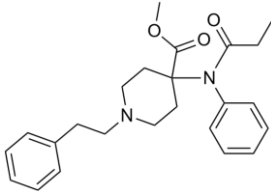
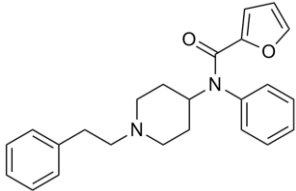
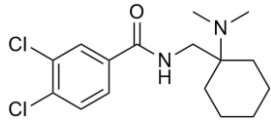
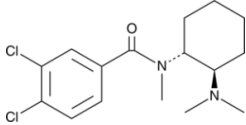
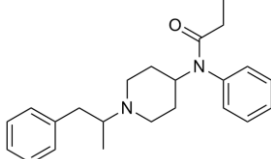
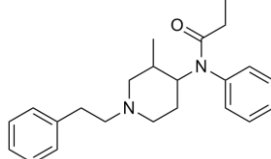
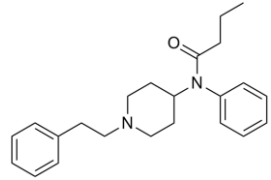
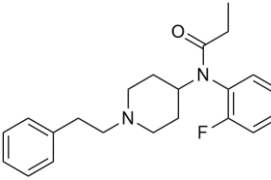
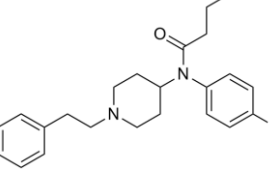
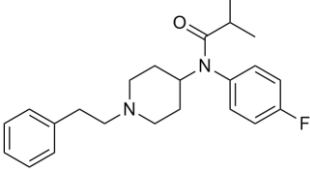
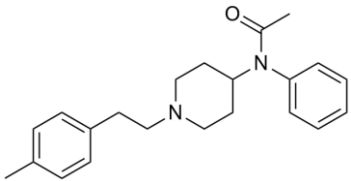
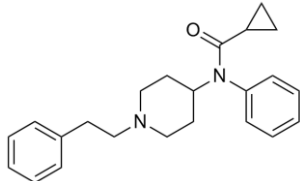
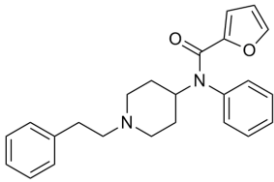
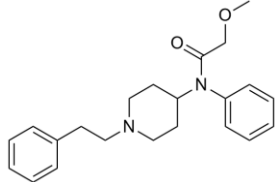
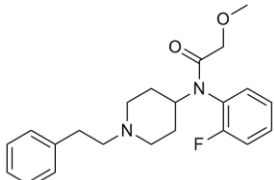
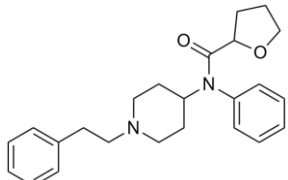
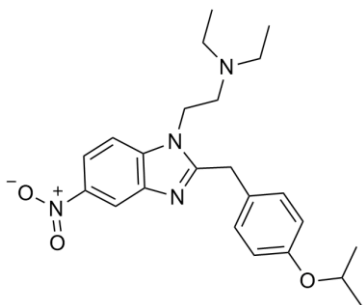
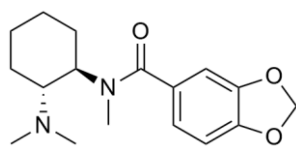
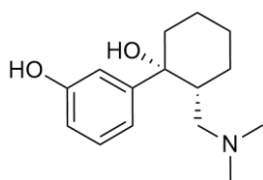
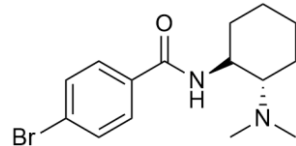
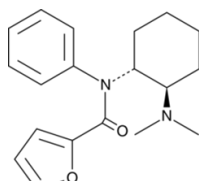
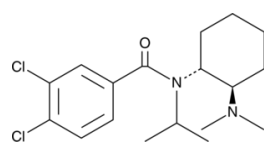
<p>Pinazepam</p> 	NB33_C	<p>Prazepam</p> 	NB34_C
<p>Temazepam</p> 	NB35_C	<p>Tetrazepam</p> 	NB36_C
<p>Triazolam</p> 	NB37_C		
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>Flubromazolam</p> 	NB38_C	<p>Clonazepam</p> 	NB39_C
<p>Diclazepam</p> 	NB40_C		

Tabela 3.6 Lista de drogas pertencentes à família dos Opióides Sintéticos divididos pelas categorias Controlo (C) e Não ilícitas (NI).

Opióides Sintéticos			
Controlo	Abreviatura	Controlo	Abreviatura
Acetil fentanil 	OS1_C	Acriloilfentanil 	OS2_C
Carfentanil 	OS3_C	Furanyl fentanil 	OS4_C
AH-7921 	OS5_C	U-47700 	OS6_C
α -metilfentanil 	OS7_C	3-metilfentanil 	OS8_C
Butirilfentanil 	OS9_C	2-Fluorofentanil 	O10_C
4-Fluorobutirifentanil 	OS11_C	4-Fluoroisobutirifentanil 	OS12_C

<p>Acetil fentanil 4-methylphenethyl</p> 	OS13_C	<p>Cyclopropilfentanil</p> 	OS14_C
<p>Furanilfentanil</p> 	OS15_C	<p>Methoxiacetilfentanil</p> 	OS16_C
<p>Ocfentanil</p> 	OS17_C	<p>Tetrahidrofuranilfentanil</p> 	OS18_C
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>Isotonitazene</p> 	OS19_NI	<p>3,4-Metilenedioxi-U-47700</p> 	OS20_NI
<p>O-Desmetiltramadol</p> 	OS21_NI	<p>Bromadoline</p> 	OS22_NI
<p>Furanil UF-17</p> 	OS23_NI	<p>Isopropil-U-47700</p> 	OS24_NI

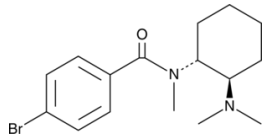
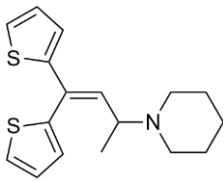
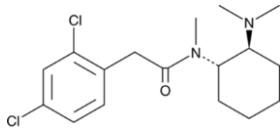
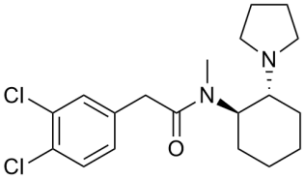
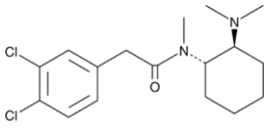
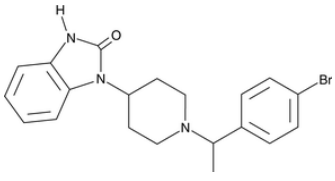
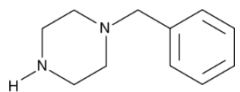
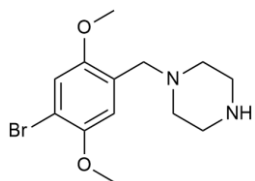
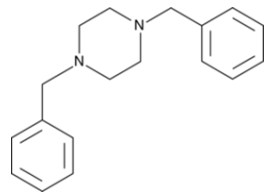
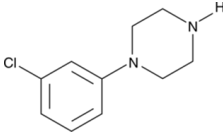
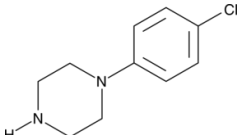
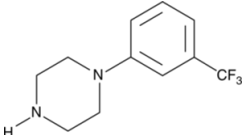
<p>N-Metil U-47931E</p> 	OS25_NI	<p>Piperidiltiambuteno</p> 	OS26_NI
<p>U-48800</p> 	OS27_NI	<p>U-50488</p> 	OS28_NI
<p>U-51754</p> 	OS29_NI	<p>Brorfine</p> 	OS30_NI

Tabela 3.7 Lista de drogas pertencentes à família das Piperazinas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

Piperazinas			
Controlo	Abreviatura	Controlo	Abreviatura
<p>BZP</p> 	Pi1_C	<p>2C-B-BZP</p> 	Pi2_C
<p>DBZP</p> 	Pi3_C	<p>mCPP</p> 	Pi4_C
<p>pCPP</p> 	Pi5_C	<p>TFMPP</p> 	Pi6_C

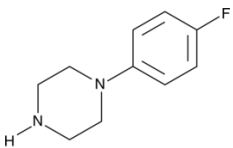
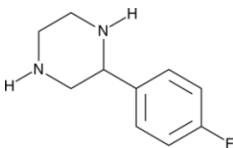
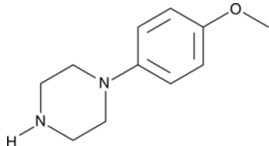
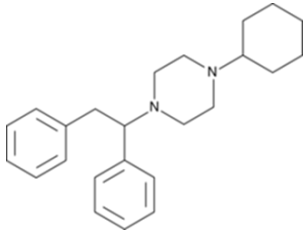
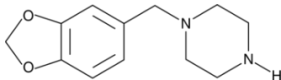
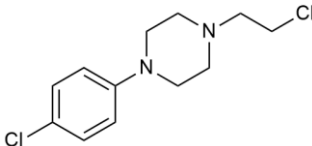
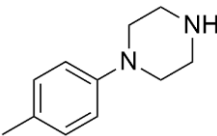
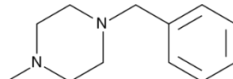
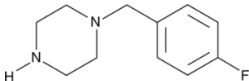
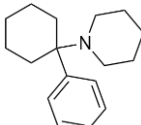
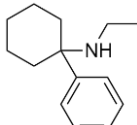
<p>1-(4-Fluorofenil)piperazina</p> 	Pi7_C	<p>2-(4-Fluorofenil)piperazina</p> 	Pi8_C
<p>MeOPP</p> 	Pi9_C	<p>MT-45</p> 	Pi10_C
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>MDBP</p> 	Pi11_NI	<p>mCPCPP</p> 	Pi12_NI
<p>MePP</p> 	Pi13_NI	<p>MBZP</p> 	Pi14_NI
<p>4-fluoro BZP</p> 	Pi15_NI		

Tabela 3.8 Lista de drogas pertencentes à família das Substâncias tipo Fenciclidinas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

Substâncias tipo Fenciclidinas			
Controlo	Abreviatura	Controlo	Abreviatura
<p>PCP</p> 	Ps1_C	<p>PCE</p> 	Ps2_C

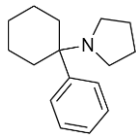
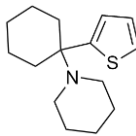
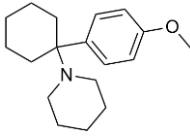
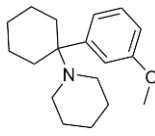
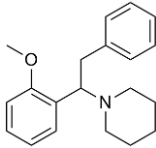
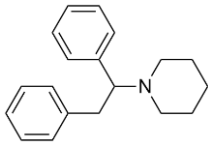
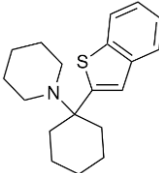
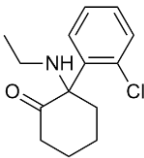
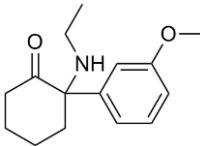
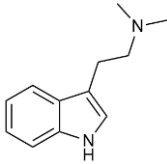
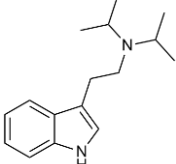
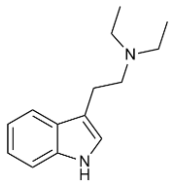
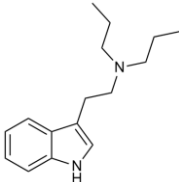
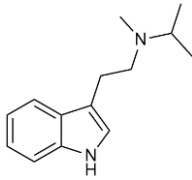
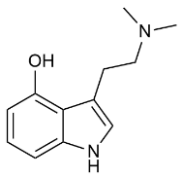
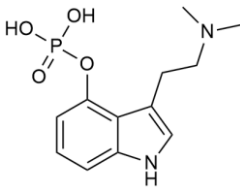
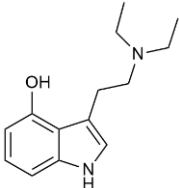
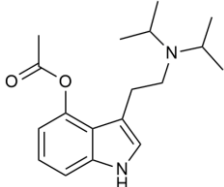
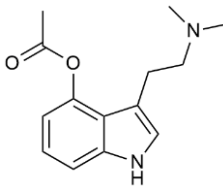
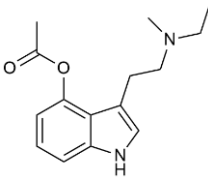
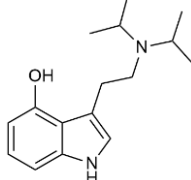
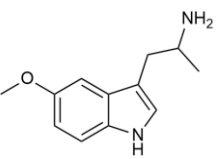
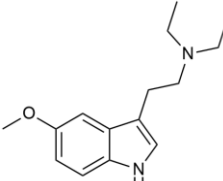
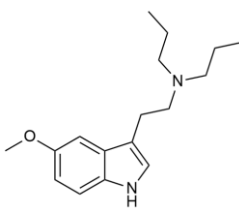
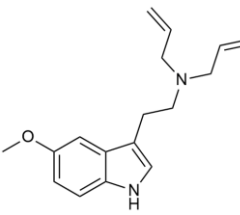
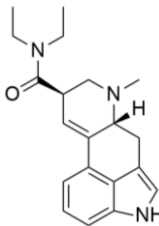
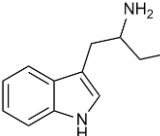
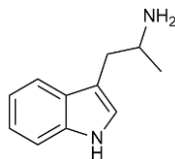
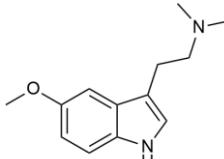
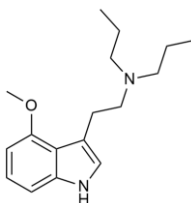
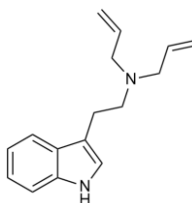
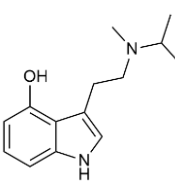
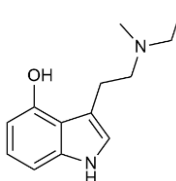
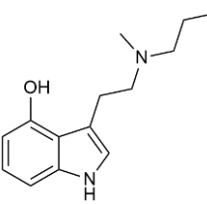
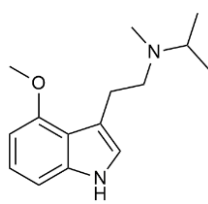
<p>PCPY</p> 	Ps3_C	<p>TCP</p> 	Ps4_C
<p>4-MeO-PCP</p> 	Ps5_C		
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>3-MeO-PCP</p> 	Ps6_NI	<p>2-methoxidifenidine</p> 	Ps7_NI
<p>Difenidine</p> 	Ps8_NI	<p>BCP</p> 	Ps9_NI
<p>2-Cl-2'-Oxo-PCE</p> 	Ps10_NI	<p>3-MeO2-Oxo-PCE</p> 	Ps11_NI

Tabela 3.9 Lista de drogas pertencentes à família das Triptaminas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

Triptaminas			
Controlo	Abreviatura	Controlo	Abreviatura
<p>DMT</p> 	Tr1_C	<p>DiPT</p> 	Tr2_C

<p>DET</p> 	Tr3_C	<p>DPT</p> 	Tr4_C
<p>MIPT</p> 	Tr5_C	<p>Psilocin</p> 	Tr6_C
<p>Psilocibin</p> 	Tr7_C	<p>4-OH-DET</p> 	Tr8_C
<p>4-AcO-DIPT</p> 	Tr9_C	<p>4-AcO-DMT</p> 	Tr10_C
<p>4-AcO-MET</p> 	Tr11_C	<p>4-OH-DIPT</p> 	Tr12_C
<p>5-MeO-AMT</p> 	Tr13_C	<p>5-MeO-DET</p> 	Tr14_C

<p>5-MeO-DPT</p> 	Tr15_C	<p>5-MeO-Dalt</p> 	Tr16_C
<p>LSD</p> 	Tr17_C	<p>Etriptamina</p> 	Tr18_C
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>AMT</p> 	Tr19_NI	<p>5-MeO-DMT</p> 	Tr20_NI
<p>4-MeO-DPT</p> 	Tr21_NI	<p>DALT</p> 	Tr22_NI
<p>4-OH-MiPT</p> 	Tr23_NI	<p>4-OH-MET</p> 	Tr24_NI
<p>4-OH-MPT</p> 	Tr25_NI	<p>4-MeO-MiPT</p> 	Tr26_NI

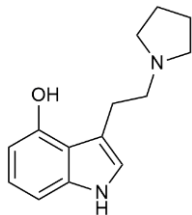
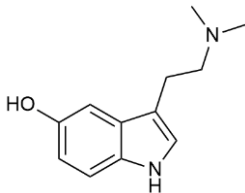
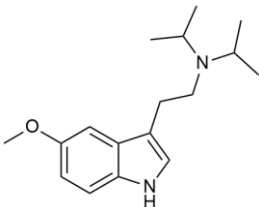
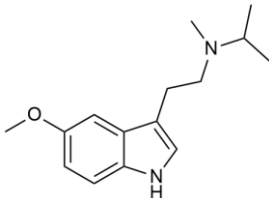
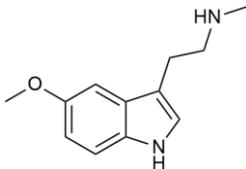
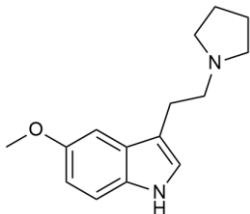
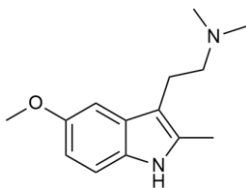
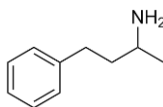
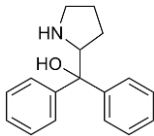
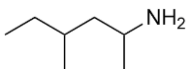
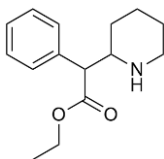
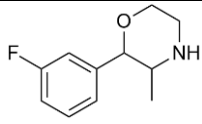
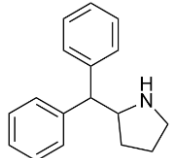
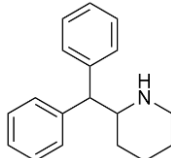
4-OH-pyr-T 	Tr27_NI	Bufotenine (5-OH-DMT) 	Tr28_NI
5-MeO-DiPT or "Foxy" 	Tr29_NI	5-MeO-MiPT or "Moxy" 	Tr30_NI
5-MeO-NMT 	Tr31_NI	5-MeO-pyr-T 	Tr32_NI
5-MeO-TMT 	Tr33_NI		

Tabela 3.10 Lista de drogas pertencentes à família das Outras drogas divididas pelas categorias Controlo (C) e Não ilícitas (NI).

Outras drogas			
Controlo	Abreviatura	Controlo	Abreviatura
3-Amino-1-phenilbutane 	OC1_C	D2PM 	OC2_C
DMAA 	OC3_C	Etilfenidate 	OC4_C

<p>LSA</p>	OC5_C	<p>MPA</p>	OC6_C
<p>Cetamina</p>	OC7_C	<p>4-Fluorotropacocaine</p>	OC8_C
<p>Dimetocaine</p>	OC9_C	<p>DMHP</p>	OC10_C
<p>4-Metilaminorex</p>	OC11_C	<p>Amineptine</p>	OC12_C
<p>Etilfenidate</p>	OC13_C	<p>Mecloqualone</p>	OC14_C
<p>Metaqualone</p>	OC15_C	<p>Metilfenidate</p>	OC16_C
<p>Fenmetrazine</p>	OC17_C		
Não ilícito	Abreviatura	Não ilícito	Abreviatura
<p>3-Fluorofenmetrazine</p>	OC18_NI	<p>4-Benzilpiperidina</p>	OC19_NI

			
Desoxi-D2PM 	OC20_NI	2-DPMP 	OC21_NI

3.2 Descritores Moleculares

Os descritores moleculares (num total de 287) referentes às moléculas descritas na secção 3.1 foram extraídos recorrendo ao pacote CDK (*Chemistry Development Kit*) do software R. Este conjunto de descritores inclui descritores Híbridos, Constitucionais, Eletrónicos, Topológicos e Geométricos. Dentro de cada categoria, os descritores dividem-se em classes tendo em consideração a característica estrutural que está a ser descrita, por exemplo, o descritor *SmallRings*, um descritor topológico, subdivide-se em *nSmallRings*, *nAromRings*, *nRingsBlock*, *nAromBlock*, *nRings3*, entre outros, e irá descrever a estrutura dos anéis presentes na molécula em estudo. Os descritores encontram-se sumariados no Anexo 1.

Os Descritores Híbridos descrevem a diversidade química da estrutura 3D de cada molécula, em termos de tamanho, forma, simetria e distribuição de átomos. Os Descritores Constitucionais, por sua vez, refletem a composição química do composto sem qualquer informação sobre a geometria ou ligação entre átomos. Os Descritores Eletrónicos contêm informação acerca da distribuição eletrónica, como por exemplo a carga atómica, energias nas orbitais, entre outros parâmetros. No que diz respeito aos Descritores Topológicos, estes caracterizam a estrutura química da molécula, tendo em conta o número e natureza das ligações e a complexidade da molécula. Por último, os Descritores Geométricos, tal como o nome indica, descrevem a molécula tendo em conta a sua geometria.

É de salientar que um descritor pode pertencer a mais do que uma categoria, como é o caso dos descritores presentes na classe *CPSA* (*Charged Partial Surface Area*), que podem ser considerados Descritores Eletrónicos e Geométricos. Na Tabela 3.11 encontram-se definidos a título de exemplo, alguns dos descritores disponíveis no pacote CDK.

Tabela 3.11 Exemplos de alguns descritores moleculares e a respetiva categoria.

Descritor	Categoria	Definição
MW	Constitucional	Peso molecular
XLogP	Constitucional	Previsão do valor para logP
nHBondDonor	Eletrónico	Número de ligações de hidrogénio dadoras
nHBondAcc	Eletrónico	Número de ligações de hidrogénio aceitadoras
topoPSA	Topológico	Área topológica da superfície polar

Capítulo 4

Discriminação entre NSP com base na
estrutura química

4. Discriminação entre NSP com base na estrutura química

Neste capítulo são apresentados os resultados mais relevantes sobre a seleção, discriminação e caracterização do perfil químico das NSP e respetivas famílias, considerando os descritores moleculares, e combinando dois métodos de análise não-supervisionada, PCA e HCA. O primeiro método, permite selecionar os descritores mais relevantes para a discriminação entre as famílias de NSP e identificar grupos-alvo de NSP para a modelação molecular. Os resultados obtidos com o HCA permitem também avaliar a similaridade estrutural das NSP e das respetivas famílias, bem como identificar grupos de NSP de controlo. A ideia principal é caracterizar e classificar de forma rápida e eficiente NSP desconhecidas e não regulamentadas tendo por base o perfil químico estabelecido com base em estruturas já conhecidas.

4.1 Seleção dos descritores moleculares

Numa fase inicial do estudo recorreu-se ao PCA com o objetivo de identificar o conjunto de descritores que melhor discrimina as 301 NSP e as respetivas famílias. Numa primeira fase, procurou-se eliminar descritores cuja informação era redundante, ou seja, foram eliminados os descritores (i) que não tinham valores atribuídos (reduzindo o número inicial de 287 para 283 descritores), (ii) que tinham variância nula (reduzindo para 232 descritores), e (iii) muito correlacionados (reduzindo para 66 descritores). No final deste pré-processamento o conjunto de dados é constituído por 301 NSP e 66 descritores moleculares (indicados na Tabela 4.1).

Tabela 4.1 Lista dos 66 descritores moleculares selecionados após o pré-processamento.

Lista dos 66 descritores moleculares					
Wnu1.unity	Weta1.unity	Weta2.unity	Weta3.unity	WD.unity	BCUTw.1l
BCUTc.1l	BCUTp.1l	nAtomLC	nBase	nRings5	TopoPSA
geomShape	PetitjeanNumber	MDEO.11	MDEO.12	MDEN.12	MDEN.13
MDEN.33	khs.dCH2	khs.dsCH	khs.sssCH	khs.aasC	khs.sNH2
khs.ssNH	khs.aaNH	khs.dsN	khs.aaN	khs.sssN	khs.ddsN
khs.aasN	khs.sOH	khs.ssO	khs.aaO	khs.sF	khs.dsssP
khs.ssS	khs.aaS	khs.sCl	khs.sBr	khs.sl	FMF
VCH.3	VCH.7	C1SP1	C2SP1	C1SP2	C3SP2

C1SP3	C2SP3	C3SP3	C4SP3	ATSm1	ATSc2
ATSc3	ATSc4	ATSc5	nHBDon	FP3A.3	RPCG
RPCS	RNCS	RPSA	MOMI.XY	LOBMIN	GRAV.6

Posteriormente, procedeu-se à análise de componentes principais, usando a matriz de correlação, para reter apenas os descritores que promovem a maior discriminação entre as moléculas e as respetivas famílias. Na tabela 4.2 encontram-se sumariados os resultados correspondentes às primeiras 25 componentes que possuem uma recuperação de, aproximadamente 83% da informação inicial.

Tabela 4.2 Valores próprios e percentagem de recuperação da informação inicial em relação ao número de componentes principais para o conjunto de dados inicial constituído por 301 moléculas e 66 descritores moleculares.

#	Valor próprio (λ_i)	Variância explicada (%)	Variância explicada cumulativa (%)
PC1	7.374	11.2	11.2
PC2	5.412	8.2	19.4
PC3	4.425	6.7	26.1
PC4	3.991	6.0	32.1
PC5	3.221	4.9	37.0
PC6	2.796	4.2	41.2
PC7	2.511	3.8	45.0
PC8	2.330	3.5	48.6
PC9	2.016	3.1	51.6
PC10	1.939	2.9	54.6
PC11	1.775	2.7	57.3
PC12	1.701	2.6	59.8
PC13	1.599	2.4	62.2
PC14	1.556	2.4	64.6
PC15	1.431	2.2	66.8
PC16	1.411	2.1	68.9
PC17	1.282	1.9	70.8
PC18	1.254	1.9	72.7

PC19	1.193	1.8	74.5
PC20	1.083	1.6	76.1
PC21	1.069	1.6	77.7
PC22	1.040	1.6	79.3
PC23	0.992	1.5	80.8
PC24	0.976	1.5	82.3
PC25	0.806	1.2	83.5

Como já mencionado na secção 2.2.3, existem 3 critérios para determinar o número de componentes principais para descrever um conjunto de dados. O critério de Kaiser diz-nos que o valor próprio (λ) tem de ser superior que 1, assim sendo, segundo este critério seriam necessárias 22 componentes principais. Já o critério de Pearson afirma que o número de componentes principais escolhido tem de apresentar cerca de 80% da informação. O número de componentes a ser escolhido seria, assim, de 23 componentes.

Através do *Screeplot* da Figura 4.1 é difícil de determinar o número de componentes principais pois não existe uma diferença significativa entre as diferentes componentes.

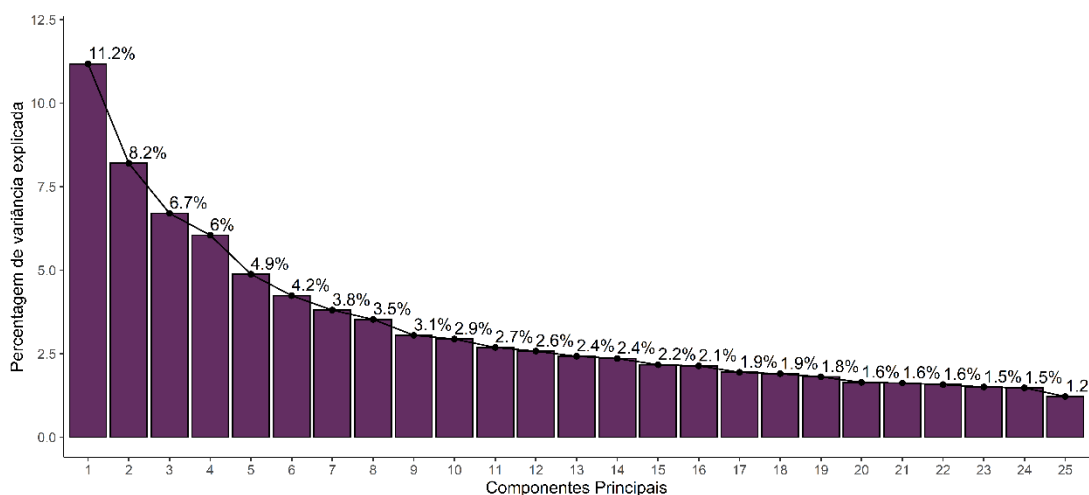


Figura 4.1 *Screeplot* para o conjunto de dados inicial constituído por 301 moléculas e 66 descritores moleculares.

De forma a assegurar a discriminação visual das moléculas representadas no novo sistema de eixos (as componentes principais), a seleção do número de descritores moleculares relevantes para a discriminação das moléculas foi realizada recorrendo ao critério descrito na referência ¹⁰³ pela expressão:

$$ndesc = 1/\sqrt{m} \quad (4.1)$$

em que m corresponde ao número total de descritores no conjunto de dados. Este critério afirma que os descritores relevantes possuem uma *loading* superior a $1/\sqrt{m}$. Neste caso o valor obtido é de 0.123. No conjunto de 66 descritores, apenas 27 deles possuem uma *loading* superior a 0.123. Estes 27 descritores foram então usados para construir um novo conjunto de dados constituídos por 301 moléculas e 27 descritores moleculares (indicados na Tabela 4.3).

Tabela 4.3 Lista dos 27 descritores moleculares selecionados após a aplicação do critério definido pela expressão 4.1.

Lista dos 27 descritores moleculares					
Weta3.unity	BCUTc.1l	nAtomLC	TopoPSA	MDEO.11	MDEO.12
MDEN.12	MDEN.13	khs.aasC	khs.sNH2	khs.ssNH	khs.dsN
khs.aaN	khs.aasN	khs.ssO	khs.sCl	FMF	VCH.7
C1SP2	C3SP2	C4SP3	ATSm1	nHBDon	FPSA.3
RPSA	MOMI.XY	GRAV.6			

Uma nova análise de PCA foi realizada sobre este conjunto reduzido. Na tabela 4.4 estão reunidos os resultados obtidos para as primeiras 11 componentes principais e reúnem cerca de 82% da variabilidade inicial.

Tabela 4.4 Valores próprios e percentagem de recuperação da informação para o conjunto de dados constituído por 301 moléculas e 27 descritores moleculares.

#	Valor próprio (λ_i)	Variância explicada (%)	Variância explicada cumulativa (%)
PC1	5.444	20.2	20.2
PC2	3.937	14.6	34.8
PC3	2.828	10.5	45.3
PC4	2.090	7.7	53.0
PC5	1.764	6.5	59.5
PC6	1.273	4.7	64.2
PC7	1.146	4.2	68.4
PC8	1.014	3.8	72.2
PC9	0.993	3.7	75.9
PC10	0.923	3.4	79.3
PC11	0.726	2.7	82.0

O critério de Kaiser indica que o número de componentes principais é de 8 e o critério de Pearson sugere 11 componentes principais.

O *Screeplot*, representado na figura 4.2, permite-nos observar que existe uma diferença mais acentuada entre componentes quando comparado com o conjunto de dados inicial, contudo a diferença que se procura ainda não é significativa para determinar o número de componentes principais através deste critério. Contudo, já foi possível obter uma maior percentagem de variância explicada para as duas primeiras componentes principais.

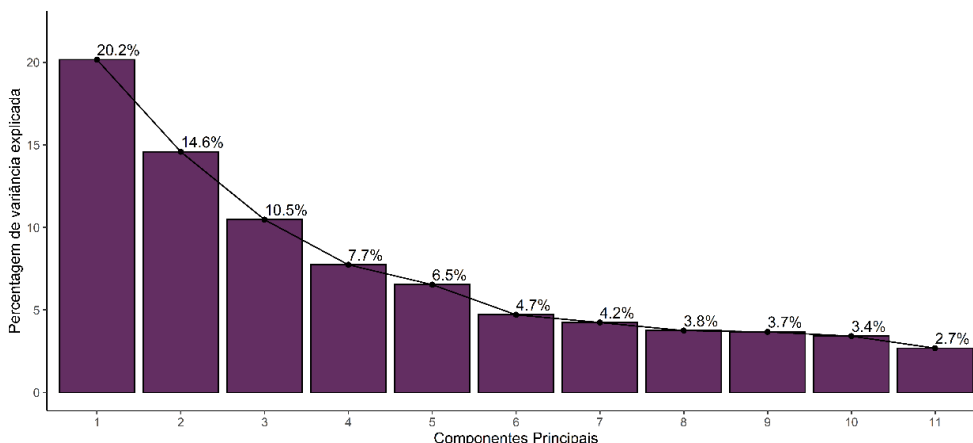


Figura 4.2 *Screeplot* para o conjunto de dados formado por 301 moléculas e 27 descritores moleculares.

No entanto, considerámos que a informação recuperada não seria suficiente para caracterizar o sistema. Neste contexto, foram encontrados dados na literatura que indicavam a existência de um conjunto de descritores que permitem fornecer indicações sobre a possível atividade biológica da molécula, sendo denominados por descritores de ADMET (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade).¹⁰⁴ Estes descritores foram adicionados ao conjunto de 27 descritores de forma a resolver melhor o sistema. O conjunto passou então a conter 53 descritores (indicados na Tabela 4.5).

Tabela 4.5 Lista dos 53 descritores moleculares após serem adicionados os descritores de ADMET.

Lista dos 53 descritores moleculares					
Weta3.unity	BCUTc.1l	nAtomLC	TopoPSA	MDEO.11	MDEO.12
MDEN.12	MDEN.13	khs.aasC	khs.sNH2	khs.ssNH	khs.dsN
khs.aaN	khs.aasN	khs.ssO	khs.sCl	FMF	VCH.7
C1SP2	C3SP2	C4SP3	ATSm1	nHBDdon	FPSA.3
RPSA	MOMI.XY	GRAV.6	MW	nRotB	ALogP
nSmallRings	nAromRings	nHBacc	TPSA	apol	naAromAtom
nAromBond	nAtom	bpol	C2SP2	C1SP3	C2SP3

C3SP3	MLogP	nRingBlock	nRings3	nRings5	nRings6
nRings7	nB	nBase	nAtomP	nAtomLAC	

Através do critério descrito pela expressão 4.1, o número de descritores foi reduzido para 43 (ver Tabela 4.6).

Tabela 4.6 Lista dos 43 descritores moleculares selecionados após a aplicação do critério definido pela expressão 4.1.

Lista dos 43 descritores moleculares					
BCUTc.1l	nAtomLC	TopoPSA	MDEO.11	MDEO.12	MDEN.12
MDEN.13	khs.sNH2	khs.ssNH	khs.dsN	khs.aaN	khs.aasN
khs.sCl	FMF	VCH.7	C1SP2	C3SP2	C4SP3
ATSm1	nHBDon	FPSA.3	MOMI.XY	MW	nSmallRings
nHBAcc	apol	naAromAtom	nAromBond	bpol	C2SP2
C1SP3	C2SP3	C3SP3	MLogP	nRingBlock	nRings3
nRings5	nRings6	nRings7	nB	nBase	nAtomP
nAtomLAC					

Na tabela 4.7 estão presentes os valores relativos às primeiras 12 componentes considerando agora o conjunto de 301 moléculas e 43 descritores moleculares.

Tabela 4.7 Valores próprios e percentagem de informação recuperada para o conjunto de dados contendo 301 moléculas e 43 descritores moleculares.

#	Valor próprio (λ_i)	Variância explicada (%)	Variância explicada cumulativa (%)
PC1	12.253	28.5	28.5
PC2	5.441	12.7	41.2
PC3	4.242	9.9	51.1
PC4	2.800	6.5	57.6
PC5	2.178	5.1	62.7
PC6	1.985	4.6	67.3
PC7	1.597	3.7	71.0
PC8	1.388	3.2	74.2
PC9	1.327	3.1	77.3
PC10	1.154	2.7	80.0

PC11	1.061	2.4	82.4
PC12	0.879	2.0	84.4

A informação que mais se destaca nesta tabela é o valor da percentagem da variância cumulativa explicada pelas duas primeiras componentes principais ter aumentado significativamente.

Relativamente ao screeplot, presente na Figura 4.3, é possível observar que a diferença entre a primeira componentes principal (PC1) com $\lambda > 12.253$ é significativa quando comparada com a segunda componente principal (PC2).

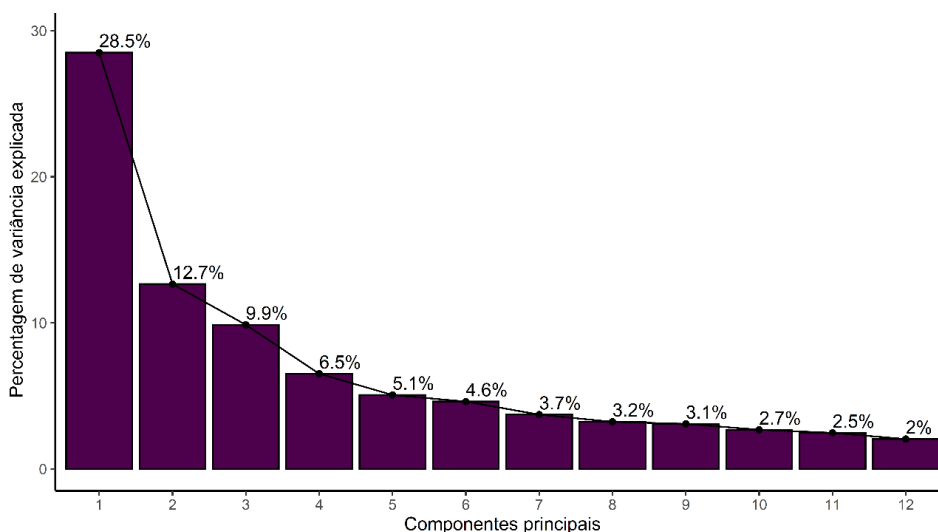


Figura 4.3 Screeplot para o conjunto de dados formado por 301 moléculas e 43 descritores.

Tendo por base a análise inicial do conjunto de dados acima descrita, foram consideradas as primeiras duas PC's para descrever o sistema. A definição dos 43 descritores que permaneceram no estudo encontra-se sumariada no Anexo 2.

A análise das *loadings* permitiu avaliar o impacto e dependência dos diferentes descritores nas componentes selecionadas, isto é, PC1 e PC2 (Figura 4.4). De forma a facilitar a visualização dos descritores, recorreu-se a uma representação gráfica individual (Figura 4.5) considerando, individualmente, cada componente principal.

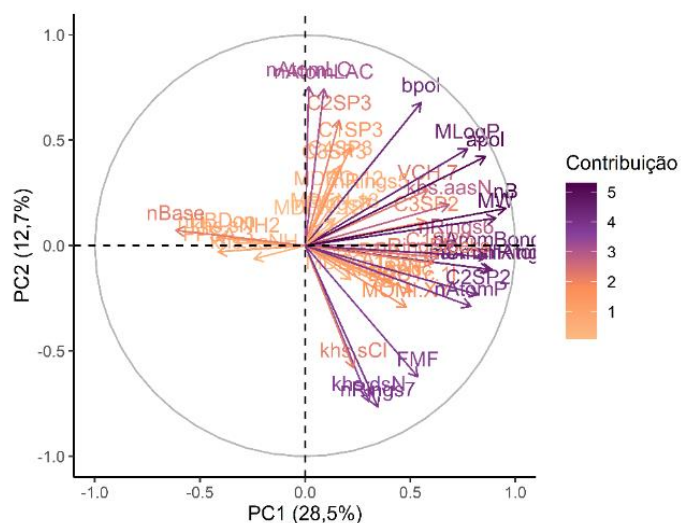


Figura 4.4 Representação gráfica da posição das variáveis relativamente às duas primeiras componentes principais.

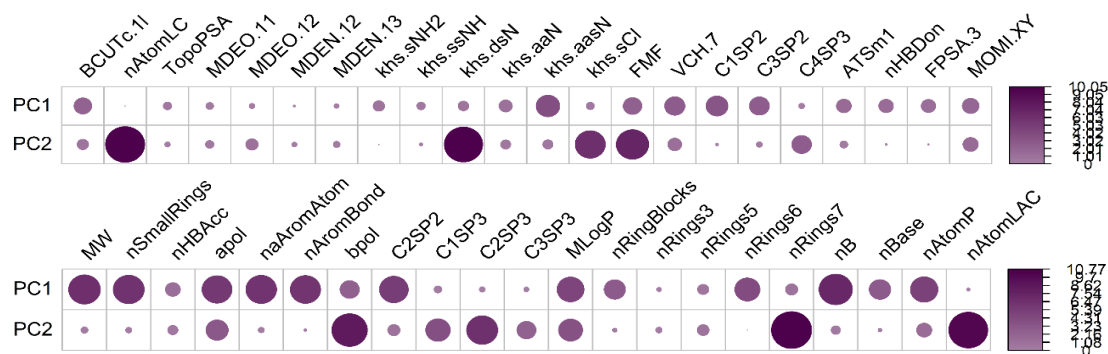


Figura 4.5 Relação entre os descritores e as duas primeiras componentes principais.

As representações gráficas presentes nas Figuras 4.4 e 4.5 são complementares e permitem avaliar o impacto que cada descritor molecular tem na PC1 e PC2. Os descritores que possuem um maior impacto na primeira componente principal são: MW, nSmallRings, apol, naAromAtom, nAromBond, C2SP2, MLogP, nRings6, nB e nAtomP; na segunda componente principal são: nAtomLC, khs.dsN, khs.sCL, FMF, bpol, C2SP3, nRings7 e nAtomLAC.

A representação na forma de biplot, Figura 4.6, facilita a visualização da posição das diferentes famílias de moléculas de NSP e do impacto dos descritores moleculares na discriminação das NSP. No biplot cada ponto é representado por uma determinada cor correspondente a uma família específica.

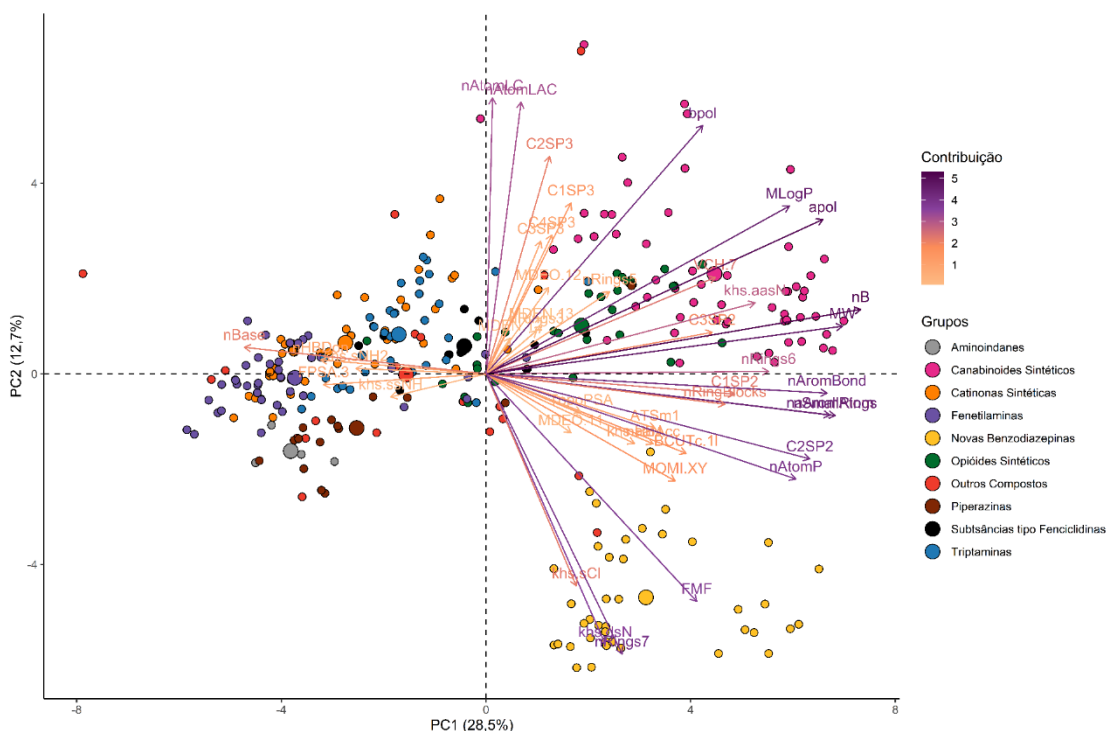


Figura 4.6 Representação na forma de biplot resultante da aplicação da análise de componentes principais sobre o conjunto de dados constituído pelas 301 moléculas e 43 descritores moleculares em função das duas primeiras componentes principais, PC1 e PC2, com uma recuperação de 41,2%.

É possível observar que quase todos os descritores se encontram no lado positivo da PC1. Por consequência, as moléculas que se situam no lado direito da PC1 irão apresentar valores mais elevados para todos esses descritores moleculares quando comparado com as moléculas que se posicionaram no lado negativo do eixo PC1. De entre os descritores que possuem um maior impacto na PC1, a maioria deles encontram-se na categoria dos descritores constitucionais. Por outro lado, na componente PC2 os descritores que detêm de uma maior contribuição para a discriminação são descritores topológicos.

Além disso, podemos ver que de uma forma geral se formam 3 grandes grupos. O primeiro é constituído quase exclusivamente por moléculas pertencentes à família das Novas Benzodiazepinas que se encontra no terceiro quadrante; outro grupo é formado na sua maioria por compostos da família dos Canabinoides Sintéticos, mas no qual também estão presentes moléculas de outras famílias, como por exemplo dos Opióides Sintéticos, todas estas moléculas posicionam-se no segundo quadrante; e, por fim entre o primeiro e quarto quadrante encontra-se uma diversidade bastante elevada de famílias, nomeadamente as Catinonas Sintéticas, Triptaminas e Fenetilaminas, entre outras. Tanto o primeiro grupo mencionado, como o segundo grupo encontram-se no lado positivo da PC1, contudo distinguem-se ao longo da segunda componente principal, PC2. As moléculas pertencentes a estes dois grupos distinguem-se principalmente pela forma como a estrutura molecular se encontra organizada. No que diz respeito ao terceiro grupo, a sobreposição de famílias indicam-nos que os grupos funcionais presentes nas diferentes estruturas são possivelmente semelhantes, possuindo assim características constitucionais e topológicas semelhantes.

4.2 Seleção das NSP baseada na similaridade estrutural

Um dos passos para atingir o objetivo deste projeto é agrupar as moléculas de acordo com a similaridade, sem que o utilizador tenha qualquer interferência no seu agrupamento. Por isso, o método escolhido foi a Análise Hierárquica de Agrupamentos (HCA), pois este método permite que as moléculas se agrupem segundo as semelhanças estruturais através de uma análise não supervisionada. Desta forma, é possível identificar grupos com características diferentes e se as moléculas se agrupam ou não com moléculas da mesma família.

Para esta análise foi utilizada a distância Euclidiana e a ligação *Ward* para representar a dissimilaridade (distância) entre as moléculas. Portanto, o dendrograma correlaciona as 301 moléculas tendo em conta os 43 descritores determinados na análise do PCA. A Figura 4.7

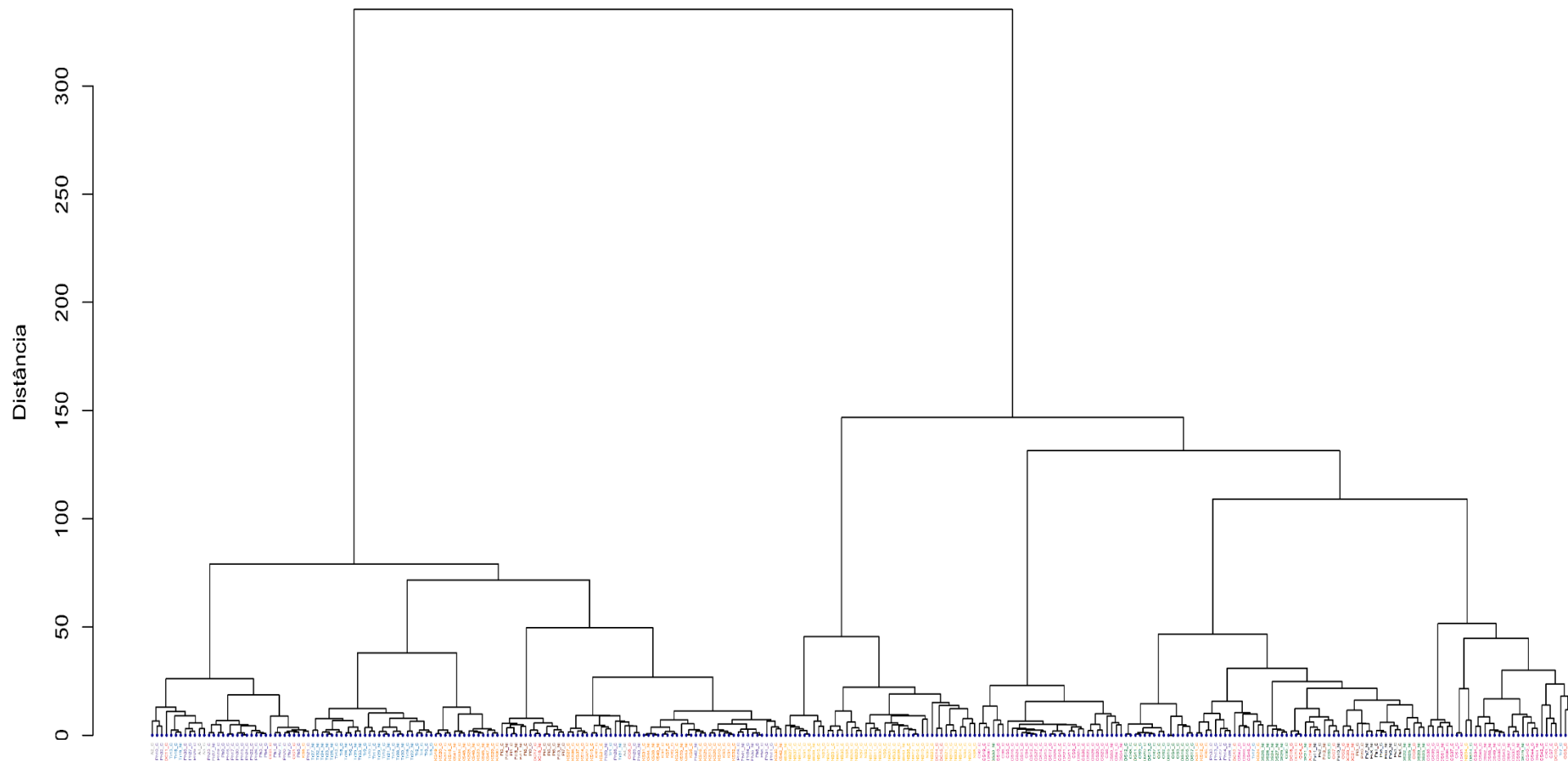


Figura 4.7 Dendrograma resultante da análise HCA usando a distância Euclidiana e método *Ward* sobre o conjunto de dados formado pelas 301 moléculas e 43 descritores moleculares. As cores das abreviaturas correspondem à família na qual as NSP estão inseridas.

corresponde ao dendrograma inicial e à semelhança da análise do PCA, as abreviaturas encontram-se representadas com cores diferentes de acordo com a família a que pertencem.

De forma a eliminar moléculas de controlo altamente correlacionadas, isto é, com o mesmo grau de similaridade em relação à molécula a catalogar, o número de NSP inicial do conjunto foi reduzido de modo a simplificar a caracterização estrutural de cada família. Especificamente, para esta redução, procurou-se os pares de moléculas Controlo-Controlo diretamente ligados que possuíam um elevado grau de similaridade e que pertenciam à mesma família, eliminando uma molécula desse par. A ideia é manter na análise apenas moléculas representativas de cada família e de cada categoria (Controlo e Não ilícitas). O número total de moléculas retiradas foi de 67 e o conjunto de dados final contém 234 moléculas e 43 descritores moleculares (ver Tabela 4.8 as NSP retiradas do estudo).

Tabela 4.8 Lista das 67 NSP eliminadas do conjunto de dados inicial.

67 NSP eliminadas do conjunto de dados inicial							
A4_C	CS1_C	CS4_C	CS10_C	CS11_C	CS14_C	CS17_C	CS18_C
CS20_C	CS24_C	CS26_C	CS28_C	CS29_C	CS30_C	CS37_C	CS39_C
CS43_C	CS45_C	CS48_C	KS2_C	KS3_C	KS7_C	KS12_C	KS14_C
KS15_C	KS16_C	KS18_C	KS20_C	KS26_C	KS29_C	KS32_C	KS36_C
NB3_C	NB13_C	NB14_C	NB16_C	NB18_C	NB19_C	NB22_C	NB29_C
NB30_C	NB31_C	NB32_C	NB33_C	NB35_C	OS2_C	OS11_C	OS12_C
OS13_C	OS15_C	OS17_C	Ph2_C	Ph4_C	Ph8_C	Ph17_C	Ph19_C
Ph20_C	Ph24_C	Ph26_C	Ph29_C	Ph32_C	Pi5_C	Ps4_C	Tr2_C
Tr5_C	Tr11_C	Tr12_C					

O novo conjunto de dados foi submetido a uma nova análise HCA, utilizando o tipo de distância e método previamente referidos. Contudo, para uma análise mais eficiente foi necessário determinar o número de grupos ótimo. Para tal utilizou-se o método *Silhouette* que permite determinar o número de grupos mais indicado para os dados em estudo tendo por base a qualidade dos agrupamentos. Para determinar o número de grupos mais indicado, este método calcula a média da qualidade para diferentes valores de k (grupos). O valor ideal para k é aquele que irá deter de uma melhor média. A representação gráfica da média em relação ao número de grupos encontra-se presente na Figura 4.8.

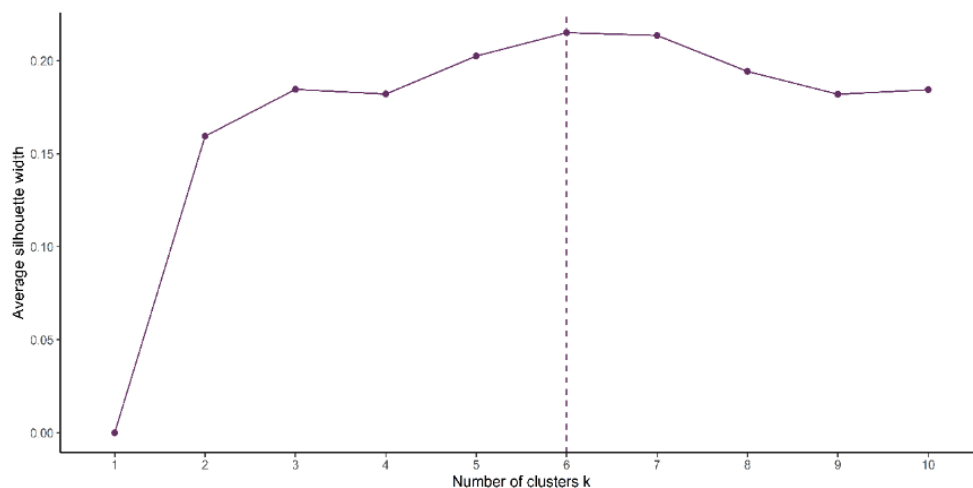


Figura 4.8 Representação gráfica do método *Silhouette* para o conjunto de dados de 234 moléculas e 43 descritores moleculares.

A Figura 4.8 sugere que o número de grupos mais adequado para o estudo em questão é de 6 grupos. Na Figura 4.9 encontra-se representado o dendrograma correspondente no qual são visíveis os 6 grupos de moléculas.

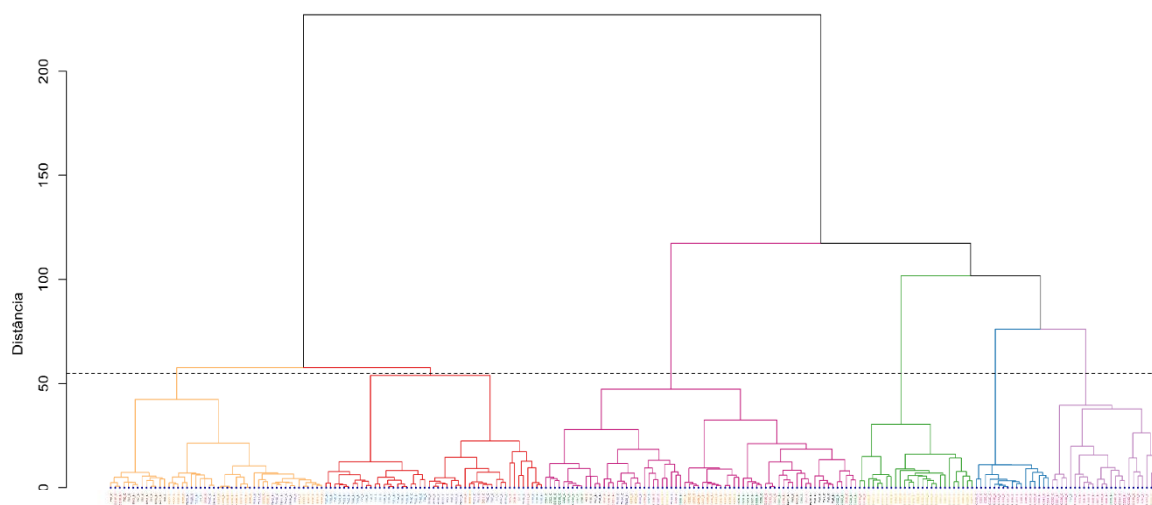


Figura 4.9 Dendrograma resultante da análise HCA usando a distância Euclidiana e método Ward sobre o conjunto de dados formado pelas 234 moléculas e os 43 descritores. As cores das abreviaturas correspondem à família na qual as NSP estão inseridas. As cores das ramificações corresponde ao grupo resultante da análise do HCA.

Analisando a Figura 4.9 em maior detalhe é possível identificar 6 grupos de moléculas e a correspondência com as 9 famílias e o seu posicionamento no plano definido pelos eixos PC1 e PC2 da Figura 4.6. Por forma a auxiliar a visualização do dendrograma apresentado na Figura 4.9, são apresentados abaixo em detalhe cada um dos 6 grupos considerados no dendrograma. A Figura 4.10 diz respeito ao primeiro grupo, representado pela cor amarela e dele fazem parte 48 moléculas de 5 famílias diferentes, nomeadamente Piperazinas,

Catinonas Sintéticas, Fenetilaminas, Triptaminas e Aminoindanos, e, também, 3 moléculas pertencentes à categoria das Outras drogas. A família mais abundante neste grupo são as Catinonas Sintéticas, contudo grande parte das moléculas pertencentes à família das Piperazinas estão presentes neste grupo.

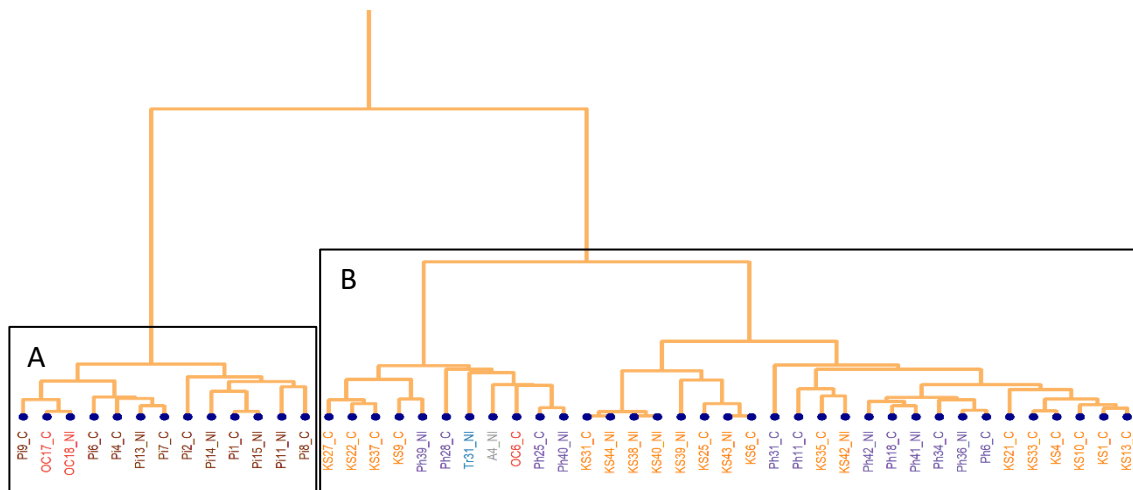


Figura 4.10 Visão detalhada do primeiro grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 48 NSP da família das Piperazinas, Catinonas Sintéticas, Fenetilaminas, Triptaminas, Aminoindanes e Outras drogas.

Nesta representação podemos ver que são criados dois subgrupos (A e B). No subgrupo A é comum a presença de um grupo fenil ligado a um anel de seis membros com pelo menos um átomo de azoto, estes anéis podem estar diretamente ligados ou por meio de CH_2 . No subgrupo B, as moléculas possuem um grupo fenil e uma amina secundária/terciária ligados por meio do grupo CH_2CH_2 . Estes dois subgrupos têm em comum a presença do grupo fenil e a presença de átomos N. Na Figura 4.11 é possível observar exemplos de moléculas diretamente ligadas e a sua semelhança estrutural.

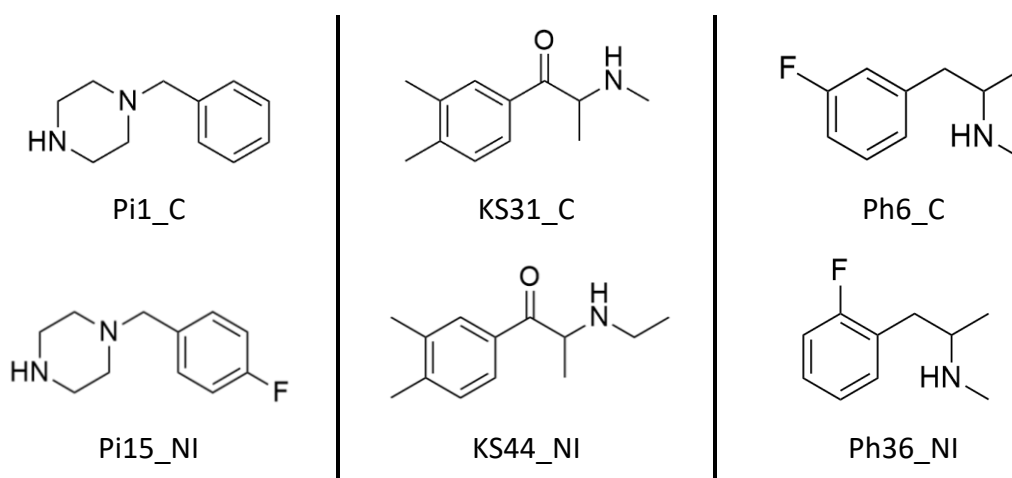


Figura 4.11 Exemplos de pares de NSP Controlo – NSP Não ilícita das famílias das Piperazinas, Catinonas Sintéticas e Fenetilaminas presentes no primeiro grupo do dendrograma da Figura 4.9.

O segundo grupo identificado no dendrograma da Figura 4.9 é apresentado com maior detalha na Figura 4.12. Este grupo é constituído por uma mistura de famílias, tais como as Triptaminas e as Fenetilaminas que são predominantes, e as Catinonas Sintéticas, Aminoindanes e Outras drogas. À semelhança do caso anterior, este também se divide em dois subgrupos (A e B).

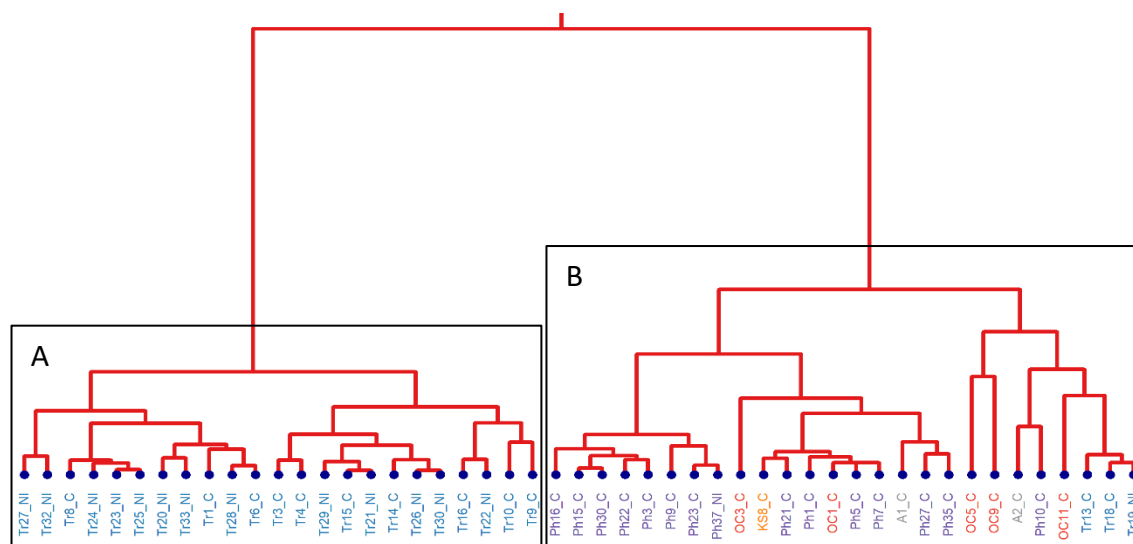


Figura 4.12 Visão detalhada do segundo grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 49 NSP da família das Triptaminas, Fenetilaminas, Catinonas Sintéticas, Aminoindanes e Outras drogas.

O primeiro subgrupo é constituído apenas por NSP da família das Triptaminas e todas elas têm na sua constituição um grupo indole e uma amina terciária ligados por uma cadeia CH_2CH_2 . O subgrupo B é comum encontrar um grupo fenil e uma amina primária na sua constituição. Ambos os grupos possuem, então, os grupos fenil e amina nas suas estruturas químicas. A Figura 4.13 diz respeito a dois pares de moléculas, um de cada subgrupo, e permite-nos observar os detalhes estruturais acima mencionados.

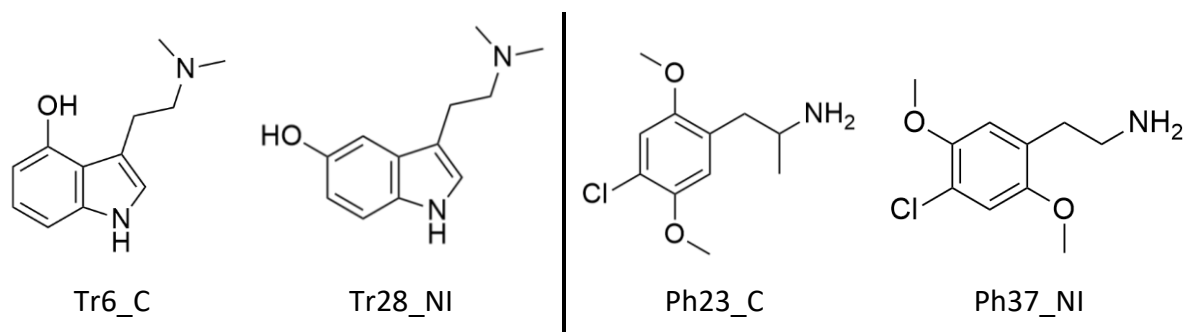


Figura 4.13 Exemplos de pares de NSP Controlo – NSP Não ilícita das famílias das Triptaminas e Fenetilaminas presentes no segundo grupo do dendrograma da Figura 4.9.

Na Figura 4.14 podemos observar com maior detalhe o terceiro grupo do dendrograma. Este é o grupo que possui o maior número de NSP (70 moléculas) e diversidade de famílias, contém NSP de todas as famílias, exceto da família dos Aminoindanos.

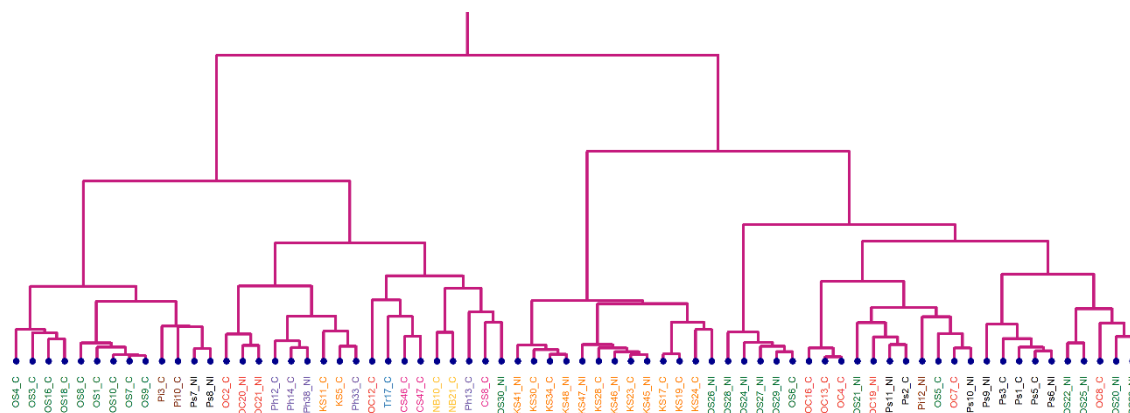


Figura 4.14 Visão detalhada do terceiro grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 70 NSP das famílias dos Opióides Sintéticos, Piperazinas, Substâncias tipo Fenciclidinas, Fenetilaminas, Catinonas Sintéticas, Triptaminas, Novas Benzodiazepinas, Canabinoides Sintéticos e Outras drogas.

Nas estruturas moleculares das NSP deste grupo é comum encontrar moléculas com pelo menos dois anéis, sendo que pelo menos um é aromático, mas também os grupos NR_3 ou NHR_2 . Na Figura 4.15 podemos observar pares de moléculas que fazem parte deste terceiro grupo do dendrograma presente na Figura 4.9.

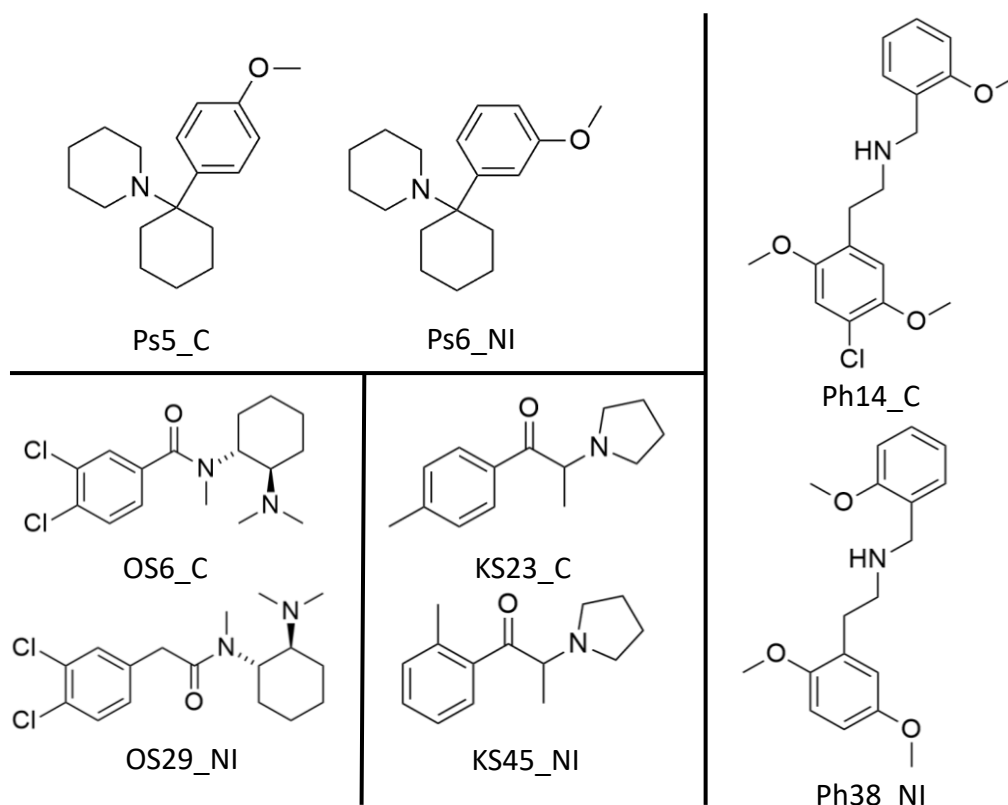


Figura 4.15 Exemplos de pares de NSP Controle – NSP Não ilícita das famílias das Substâncias tipo Fenciclidinas, Opióides Sintéticos, Catinonas Sintéticas e Fenetilaminas presentes no terceiro grupo do dendrograma da Figura 4.9.

Nos grupos presentes nas Figuras 4.10, 4.12 e 4.14 observa-se uma mistura significativa de famílias, o que era de esperar tendo em consideração os resultados obtidos na análise PCA apresentada na Figura 4.6. Estas NSP dizem respeito ao terceiro grupo mencionado na secção anterior e são discriminados pelos mesmo descritores. Com isto, é normal que as NSP possuam semelhanças estruturais entre os diferentes grupos.

O quarto grupo (Figura 4.16) é formado maioritariamente por NSP da família das Novas Benzodiazepinas e quase todas as NSP desta família se encontram neste grupo. Nas suas estruturas químicas, é comum encontrar anéis de sete membros fundidos com um benzeno ou tiofeno. O anel de sete membros pode ter 1 ou 2 átomos de azoto, bem como um grupo fenilo ligado. As NSP presentes neste grupo têm ainda átomos de halogéneos na sua composição. Como é possível observar na Figura 4.17 onde se

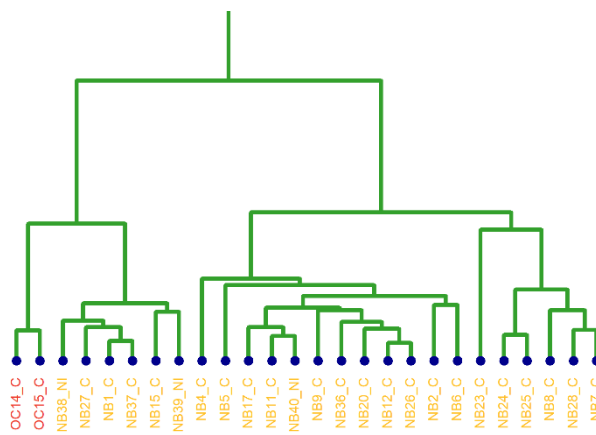


Figura 4.16 Visão detalhada do quarto grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 26 NSP das famílias dos Novas Benzodiazepinas e Outras drogas.

apresentam a título de exemplo duas moléculas pertencentes ao quarto grupo do dendrograma que demonstram.

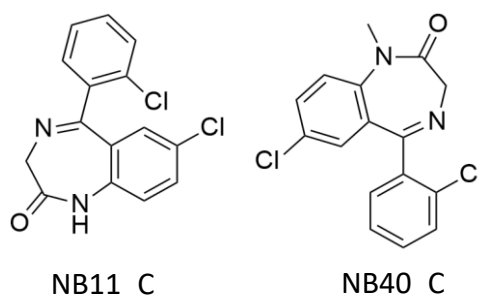


Figura 4.17 Exemplos de um par Controlo – NSP Não ilícita da família das Novas Benzodiazepinas presente no quarto grupo do dendrograma da Figura 4.9.

O quinto grupo presente no dendrograma da Figura 4.9 é apresentado com maior detalhe na Figura 4.18 é o grupo de menor dimensão, com apenas 18 NSP. Todas as NSP deste grupo fazem parte da família dos Canabinoides Sintéticos. A estrutura química dos compostos pertencentes a este grupo é muito

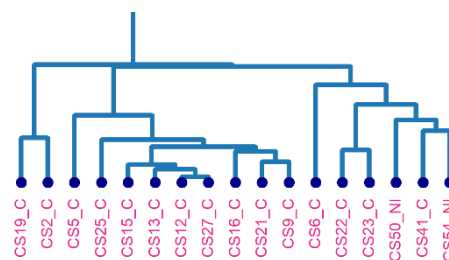


Figura 4.18 Aproximação ao quinto grupo do dendrograma presente na Figura 4.9. Neste grupo estão presentes 18 NSP da famílias dos Canabinoides Sintéticos.

semelhante. As moléculas possuem na sua constituição um grupo cetona ligado a um grupo indole. Contém pelo menos mais um anel aromático e uma cadeia lateral alifática ou um anel ligado ao átomo de azoto do índole. São apresentados a título de exemplo na Figura 4.19 a estrutura de duas moléculas presente neste quinto grupo.

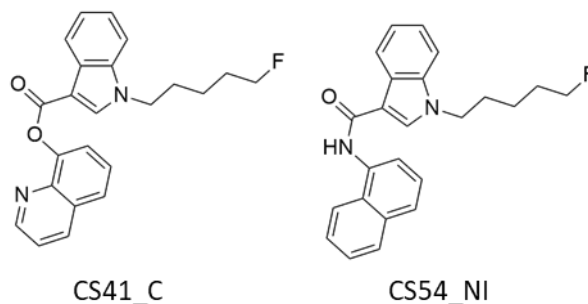


Figura 4.19 Exemplo de um par de NSP Controlo – NSP Não ilícita da família dos Canabinoides Sintéticos presente no quinto grupo do dendrograma da Figura 4.9.

Por último, o sexto grupo do dendrograma da Figura 4.6, apresentado com maior detalhe na Figura 4.20, possui 24 NSP de cinco famílias diferentes, sendo os Canabinoides Sintéticos os mais representativos, seguindo-se dos Opióides Sintéticos, Triptaminas, Novas Benzodiazepinas e Outras drogas. A estrutura molecular dos compostos pertencentes a este grupo é constituída por pelos menos um grupo fenilo. É, também, comum encontrar os grupos carbonilo, aminas e cadeias alifáticas.

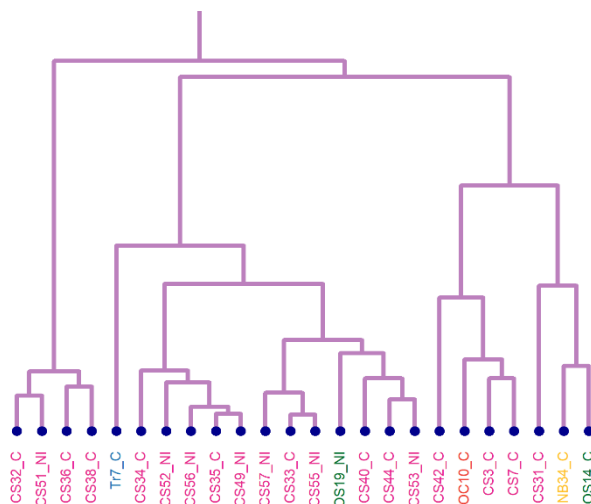


Figura 4.20 Visão detalhada do sexto grupo do dendrograma da Figura 4.9. Neste grupo estão presentes 24 NSP das famílias dos Canabinoides Sintéticos, Opióides Sintéticos, Triptaminas, Novas Benzodiazepinas e Outras drogas.

Na Figura 4.21 estão representadas as estruturas de duas moléculas, uma NSP de controlo e outra não regulamentada da família dos Canabinoides Sintéticos presente no último grupo do dendrograma da Figura 4.9.

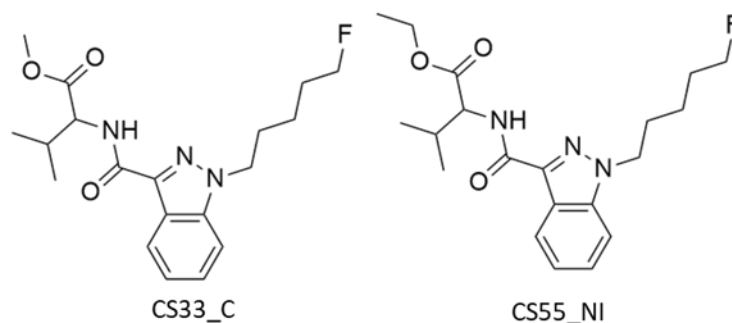


Figura 4.21 Exemplo de um par NSP Controlo – NSP Não ilícita da família dos Canabinoides Sintéticos presente no sexto grupo do dendrograma da Figura 4.9.

Tendo em consideração o conjunto de dados final, que contém 234 moléculas, procurou-se identificar quais dos 43 descritores têm maior peso na discriminação entre famílias (Figura 4.22). É de notar que as Catinonas Sintéticas e os Canabinoides Sintéticos se discriminam essencialmente pelos descritores moleculares predominantes da PC1, nomeadamente MW, MLogP, nB e apol. De acordo com este conjunto de descritores moleculares, podemos afirmar que os Canabinoides Sintéticos possuem maior peso molecular (MW), maior número de ligações (nB), maior lipofilicidade (MLogP) e maiores valores para as somas das polarizabilidades atómicas (apol) do que as Catinonas Sintéticas. Por outro lado, os Canabinoides Sintéticos e as Novas Benzodiazepinas discriminam-se, principalmente, pelos descritores mais predominantes da PC2, mais especificamente o FMF, Rings7, khs.dsN, nAtomLC e nAtomLAC. As NSP da família das Novas Benzodiazepinas apresentam uma complexidade estrutural maior (FMF), possuem anéis de 7 membros (Rings7) e fragmentos $= N - (khs.dsN)$. Contudo, nas NSP da família dos Canabinoides Sintéticos é mais comum encontrar cadeias laterais com um número maior de átomos (nAtomLC), bem como cadeias alifáticas (nAtomLAC), do que na família das Novas Benzodiazepinas. Assim, é possível verificar que os Canabinoides Sintéticos se encontram bem discriminados pelo conjunto de descritores moleculares.

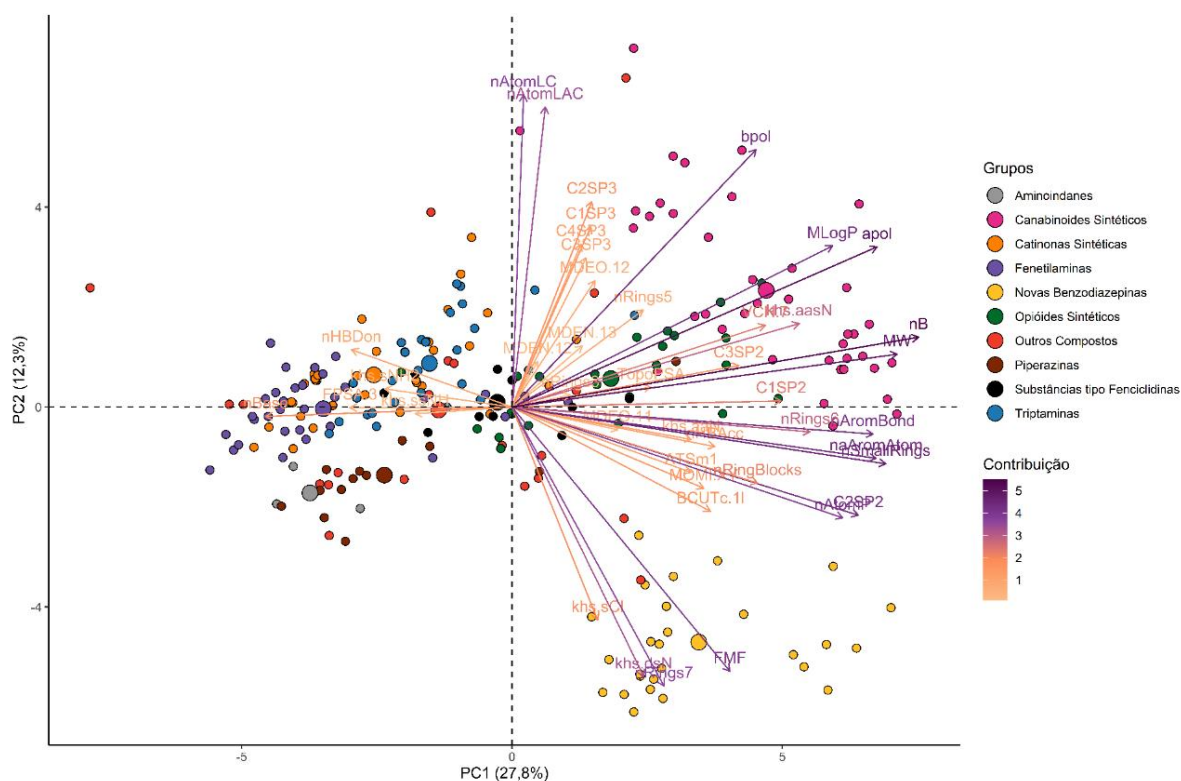


Figura 4.22 Representação na forma de biplot resultante da aplicação da análise de componentes principais sobre o conjunto de dados constituído pelas 234 moléculas e 43 descritores moleculares em função das duas primeiras componentes principais, PC1 e PC2, com 27,8% e 12,3% de variabilidade recuperada.

Tendo em conta os resultados obtidos por aplicação dos métodos PCA e HCA, procurou-se validar os agrupamentos obtidos estudando a nível molecular, o comportamento de substâncias incluídas num mesmo grupo e em grupos distintos, por forma a verificar se a similaridade/dissimilaridade obtida se traduz em comportamentos físico-químicos e biológicos diferentes. Assim, foram selecionadas duas moléculas da família dos Canabinoides Sintéticos, mais concretamente CS33_C e CS55_NI. Estas duas moléculas representam um par de moléculas Controlo - Não ilícitas da mesma família e apresentam uma similaridade muito elevada, de acordo com a análise HCA (Figura 4.20). Foi ainda selecionada a molécula CS12_C pertencente à mesma família, mas cuja análise HCA/PCA posicionou num grupo diferente (Grupo 5, Figura 4.18). É expectável que pares de moléculas Controlo - Não ilícitas colocadas num mesmo grupo apresentem comportamento idêntico e, por outro lado, que moléculas posicionadas em grupos distintos apresentem perfis de interação e características diferentes.

O estudo do comportamento destas 3 moléculas a nível molecular, nomeadamente da sua interação com o recetor alvo desta família e com uma membrana modelo, foram conduzidos recorrendo ao *Docking Molecular* e à Simulação por Dinâmica Molecular.

Capítulo 5

Detalhes das simulações moleculares

5. Detalhes das simulações moleculares

Neste capítulo serão apresentados os detalhes computacionais referentes às simulações moleculares por *docking* molecular e DM.

5.1 Dinâmica Molecular

As simulações por DM foram realizadas com o intuito de avaliar a interação das NSP, nomeadamente CS12_C (JWH-018), CS33_C (5F-AMB-PINACA) e CS55_NI (5F-EBM-PINACA), cujas estruturas se apresentam na Figura 5.1, com uma bicamada lipídica modelo. A membrana modelo utilizada neste estudo é uma bicamada fosfolipídica hidratada constituída por 128 moléculas de POPC (1-palmitoil-2-oleoil-sn-glicero-3-fosfocolina). Apesar de ser um modelo simplista, esta bicamada tem sido utilizada como modelo para mimetizar a barreira hematoencefálica.¹⁰⁵

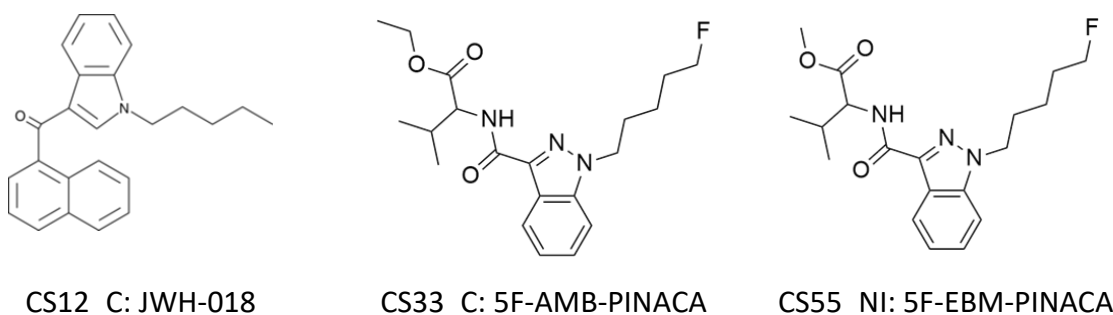


Figura 5.1 Estrutura molecular das NSP selecionadas para o estudo a nível molecular por simulação computacional. As NSP selecionadas pertencem à família dos Canabinoides Sintéticos.

Foram considerados 4 sistemas, um sistema de referência constituído unicamente pela bicamada lipídica e mais 3 sistemas constituídos respetivamente pela bicamada lipídica e cada uma das 3 moléculas de NSP selecionadas. A membrana modelo é constituída por uma bicamada hidratada, contendo 128 lípidos de POPC distribuídos igualmente pelas duas camadas e 2460 moléculas de água. A topologia da membrana foi obtida através do site do Grupo de Biocomputação da Universidade de Calgary.¹⁰⁶ A topologia das NSP foi gerada utilizando a plataforma *Automated Topology Builder* (ATB).¹⁰⁷

Para o passo da hidratação e solvatação da membrana, foi utilizado o modelo de carga pontual simples (SPC) e os parâmetros lipídicos de átomos unidos de Berder e colaboradores.¹⁰⁸ Nas simulações foi usado o campo de forças GROMOS 54a7, introduzindo as correções de Berger para os lípidos.¹⁰⁸ As NSP foram introduzidas individualmente na membrana, obtendo, tal como já mencionado, um total de 4 sistemas: sistema de referência

(bicamada de POPC), sistema 1 (bicamada de POPC e 1 molécula do canabinoide CS12_C), sistema 2 (bicamada de POPC e 1 molécula do canabinoide CS33_C) e sistema 3 (bicamada de POPC e 1 molécula do canabinoide CS55_NI). As ligações foram limitadas aos seus comprimentos de equilíbrio com o algoritmo de limitação linear (LINCS). Para as interações Lennard-Jones, foi aplicado um *cut-off* a 1-4 nm, enquanto as interações eletrostáticas de longo alcance foram tratadas com o método PME (em inglês *Particle Mesh Ewald*). Todas as simulações foram realizadas no *ensemble* isotérmico-isobárico (NPT), onde são mantidas constantes o número de partículas, a pressão e a temperatura, sob condições periódicas de fronteira. O intervalo de tempo utilizado para a resolução das equações de movimento foi definido como 2.0 fs. As interações não ligantes foram calculadas tendo por base uma lista de átomos vizinhos, que foi atualizada a cada 10 ps. Todos os sistemas foram simulados durante 200 ns. Para manter a temperatura a 323 K e a pressão a 1 bar ao longo da simulação, foram utilizados o termostato Nose-Hoover e o baróstato Parrinello-Rahman, respetivamente.

Todas as análises foram efetuadas nos últimos 20 ns da simulação, por forma a garantir que os sistemas estavam devidamente equilibrados. Os parâmetros avaliados foram os perfis de densidade ao longo da bicamada, o deslocamento quadrático médio (MSD) e o parâmetro de ordem de deutério. Configurações representativas de cada sistema foram extraídas utilizando o software VMD (versão 1.9.4a48).

5.2 Docking Molecular

A simulação por *docking* molecular foi utilizada para avaliar a afinidade das 3 moléculas de canabinoides selecionados com o seu alvo preferencial. Os canabinoides sintéticos são produzidos com o propósito de reproduzir efeitos semelhantes ao Δ^9 -THC, pelo que o recetor biológico das novas moléculas é preferencialmente o mesmo do Δ^9 -THC. Os alvos preferenciais destas moléculas são o recetor endocanabinoide tipo 1 (CB₁) e o recetor endocanabinoide tipo 2 (CB₂). Contudo, alguns estudos reportam que os canabinoides sintéticos possuem uma afinidade maior para o recetor CB₁.¹⁰⁹ Este recetor encontra-se no cérebro e sistema nervoso central, regulando uma variedade de funções e comportamentos cerebrais, como a dor e a memória. Além disso, também é possível encontrar este recetor em órgãos e tecidos periféricos, onde demonstrou desempenhar funções no metabolismo energético e no controlo do apetite.¹¹⁰

A estrutura 3D deste recetor foi extraída do Protein Data Bank (código 5XRA).¹¹¹ O primeiro passo da simulação consistiu em remover o ligando flavin mononucleotide (FMN), o colesterol, o ácido oleico, o (2R)-2,3-dihidroxipropil(9Z)-octadec-9-enoato e o dietilenoglicol, com o objetivo de reter apenas as coordenadas do recetor CB₁. A molécula de cada NSP estudada foi posicionada na caixa de simulação usando dados recolhidos na literatura sobre os resíduos mais ativos neste recetor para esta família de NSP e cujos trabalhos e respetiva informação relevante se encontra compilada na Tabela 5.1.

Tabela 5.1 Principais publicações identificando os resíduos do recetor CB1 responsáveis pela interação com moléculas da família dos Canabinoides Sintéticos.

Publicação	Resíduos identificados
1- Exploring the Ligand Efficacy of Cannabinoid Receptor 1 (CB ₁) using Molecular Dynamics Simulations ¹¹²	Phe102; Met103; Val105; Gly166; Ph170; Phe174; Phe177; Leu193; Val196; Phe200; Phe268; Phe379; Trp356; Ala380; Ser383; Met384;
2- Study of the effectiveness of various cannabinoid receptor 1 (CB ₁) agonists using molecular docking and molecular dynamics modeling ¹¹³	Phe10; Phe72; Ser75; Phe76; Phe79; Phe91; Leu95; Val98; Thr99; Phe102; Ile169; Phe170; Trp181; Met265; Lys278; Phe281; Ala282; Ser285; Cys288
3- Crystal structures of agonist-bound human cannabinoid receptor CB ₁ ¹¹¹	Phe108; Phe170; Ser173; Phe174; Phe177; His178; Thr187; Phe189; Lys192; Leu193; Val196; Phe200; Ile267; Phe268; Pro269; Ile271; Tyr275; Leu276; Trp279; Met363; Trp356; Leu359; Phe379; Ala380; Ser383; Cys386
4- Allosteric and orthosteric pharmacology of cannabidiol and cannabidiol-dimethylheptyl at the type 1 and type 2 cannabinoid receptors ¹¹⁴	Phe170; Phe174; Phe177; Leu193; Phe379; Ser383

<p>5- Cannabinoid CB₂/CB₁ Selectivity. Receptor Modeling and Automated Docking Analysis¹¹⁵</p>	<p>Cys107; Phe108; Val110; Asn112; Val168; Ile169; Phe170; Phe189; Leu193; Gly194; Val196; Thr197; Phe200; Met265; Pro269; Trp279; Trp356; Phe379; Ser383</p>
<p>6- Design, Synthesis, Binding and Docking-Based 3D-QSAR Studies of 2-Pyridylbenzimidazoles - A New Family of High Affinity CB₁ Cannabinoid Ligands¹¹⁶</p>	<p>Phe170; Phe174; Lys192; Leu193; Thr197; Phe200; Ile267; Tyr275; Leu359; Met363; Phe379; Ala380</p>
<p>7- Understanding virtual solvent through large-scale ligand discovery¹¹⁷</p>	<p>His178; Thr201; Trp279; Ser383</p>
<p>8- Structure of an allosteric modulator bound to the CB₁ cannabinoid receptor¹¹⁰</p>	<p>Phe170; Ser173; Phe174; Phe177; His178; Phe189; Lys192; Leu193; Val196; Thr197; Phe268; Ile271; Tyr275; Leu276; Trp279; Trp356; Met363; Phe379; Ser383</p>
<p>9- A computational study on cannabinoid receptors and potent bioactive cannabinoid ligands: homology modeling, docking, de novo drug design and molecular dynamics analysis¹¹⁸</p>	<p>Leu193; Phe200; Thr201; Ile247; Pro251; Thr283; Trp356; Leu360; Val387</p>
<p>10- Ligand-specific homology modeling of human cannabinoid (CB₁) receptor¹¹⁹</p>	<p>Phe174; Phe177; Thr187; Phe189; Lys192; Phe200; Tyr275; Trp279; Cys355; Trp356; Leu359; Leu360; Phe379; Phe381; Ser383 ; Met384; Cys386</p>
<p>11- Crystal Structure of the Human Cannabinoid Receptor CB₁¹²⁰</p>	<p>Phe102; Met103; Val105; Ile119; Gly166 ; Ser167; Phe170; Phe174; Leu193; Val196; Thr197; Phe268; Trp356; Leu359; Met363; Phe379; Ala380; Ser383; Met384; Cys386; Val387</p>

12- Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-Gi Complex Structures ¹²¹	Phe177; Leu193; Phe200; Phe268; Tyr275; Trp279; Trp356; Phe379; Ser383
13- Assessment of select synthetic cannabinoid receptor agonist bias and selectivity between the type 1 and type 2 cannabinoid receptor ¹²²	Phe170; Ser173; Phe177; His178; Arg182; Phe189; Phe200; Phe269; Trp279; Leu376; Ser383
14- Study of an Active-State CB ₁ Receptor Model and JWH Compound Interactions to Predict New Emerging Synthetic Cannabinoids ¹⁰⁹	Asp163; Ph3170; Phe173; Phe174; Phe177; Lys192; Leu193; Val196; Thr197; Ser199; Phe200; Ile271; Asp272; Met277; Met278; Trp279; Leu286; Val291; Trp356; Leu359; Leu360; Met363; Phe379; Ala380; Ser383; Cys386

Atendendo à informação recolhida com base na Tabela 5.1 foram identificados os resíduos mais ativos para esta classe de moléculas tendo sido selecionados os Phe170, Phe174, Phe177, Leu193, Phe200, Phe268, Trp279, Trp356, Phe379 e Ser383 de acordo com a distribuição descrita na Tabela A.7 do Anexo 3. Assim, para cada um dos 3 sistemas, cada NSP foi posicionada na caixa de simulação na proximidade destes resíduos, considerando o centroide obtido a partir das respetivas coordenadas (ver Tabela A.8 do Anexo 4). A caixa de simulação foi gerada de modo a incluir os resíduos selecionados, com o centro definido a -41.071, -162.866 e 305.740 Å e com a dimensão de 90, 80 e 76 pontos nas direções x, y e z respetivamente, com um espaçamento de 0,5 Å. Posteriormente, usando o *Genetic Lamarkian Algorithm* foram efetuados 100 passos de pesquisa no *AutoDock*. Todas as NSP foram submetidas a este processo três vezes. As conformações com menor energia de ligação foram analisadas utilizando o servidor *Protein Plus* (<https://proteins.plus/>) e o software *Discovery Studio Visualizer*.

Capítulo 6

Caracterização dos padrões de
interação

6. Caracterização dos padrões de interação

Neste capítulo são apresentados os resultados mais relevantes sobre o comportamento ao nível molecular das NSP selecionadas na secção anterior, na presença do recetor endocanabinoide tipo 1 (CB₁) e de uma membrana de 1-palmitoil-2-oleoil-sn-glicero-3-fosfocolina, usada como modelo da barreira hematoencefálica (BBB), caracterizado por *docking* molecular e DM.

6.1 Interação entre as NSP e uma membrana modelo

De modo explorar ao nível molecular, e por DM, a interação entre as moléculas CS12_C (JWH-018), CS33_C (5F-AMB-PINACA) e CS55_NI (5F-EBM-PINACA) e uma membrana fosfolipídica simples, usada como modelo para a BBB, as moléculas foram introduzidas, individualmente, numa bicamada constituída por 128 lípidos de POPC (Figura 6.1), previamente solvatada e equilibrada. Este modelo tem sido reconhecido como um modelo simplista adequado para mimetizar a BBB.¹⁰⁵

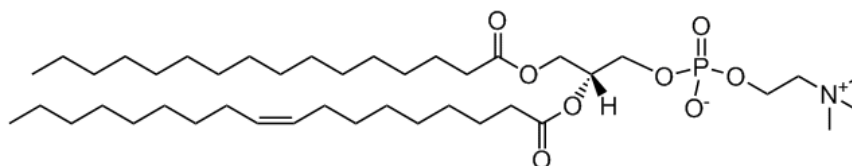


Figura 6.1 Estrutura química do lípido 1-palmitoil-2-oleoil-sn-glicero-3-fosfocolina (POPC) constituinte da membrana modelo utilizada no estudo por DM.

O impacto da incorporação de uma molécula de CS12_C, CS33_C e CS55_NI nas características estruturais da bicamada de POPC foi avaliada em primeiro lugar através dos perfis de densidade representados da Figura 6.2. Esses perfis de densidade representam o posicionamento relativo dos grupos de átomos de referência pertencentes aos fosfolípidos e às NSP em relação ao centro da bicamada, definida como região entre as duas camadas da membrana. Para as moléculas de POPC, foram selecionados os átomos P e N dos grupos polares localizados nas cabeças dos lípidos (região hidrofílica) e os grupos metilo terminais localizados nas cadeias alifáticas, hidrofóbicas. As NSP foram representadas pelo respetivo centro de massa.

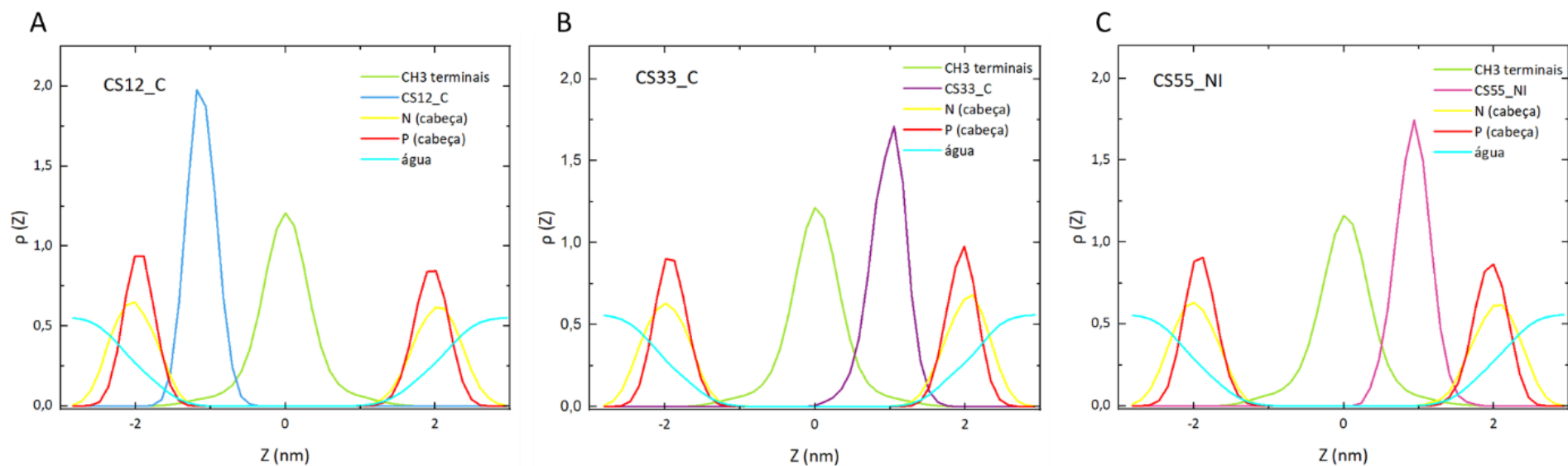


Figura 6.2 Perfis de densidade de (A) CS12_C (JWH-018), (B) CS33_C (5F-AMB-PINACA), (C) CS55_NI (5F-EBM-PINACA) através da bicamada de POPC (1 NSP:128 POPC). As moléculas de POPC foram identificadas pelos átomos P (vermelho) e N (amarelo) dos grupos polares e pelos átomos dos grupos metilo terminais (verde). O solvente (água) encontra-se representado pelo perfil azul-claro. As NSP foram representadas pelos respetivos centros de massa (azul, roxo e cor-de-rosa, respetivamente). De notar que a coordenada Z representa o corte longitudinal da membrana.

Os perfis de densidade dos compostos em estudo localizam-se em zonas bastante próximas, contudo é possível observar que os perfis correspondentes a CS33_C e CS55_NI são idênticos. Relativamente à molécula CS12_C observa-se uma menor penetração na zona central da bicamada do que as duas outras moléculas de NSP. Ou seja, ainda que pouco expressiva, verifica-se uma menor afinidade desta molécula, pela zona mais hidrofóbica da bicamada lipídica. Na Figura 6.3 é possível observar configurações típicas dos compostos na membrana modelo obtidas na simulação de Dinâmica Molecular.

Foi ainda avaliada a perturbação que cada uma das moléculas indiz na estrutura e organização da bicamada lipídica, através dos perfis do parâmetro de ordem de deutério (S_{cd}) e do Deslocamento Quadrático Médio (MSD) que se encontram representados na Figura 6.4. Nestas análises é usado como sistema de referência uma bicamada de POPC solvatada, na ausência de qualquer molécula de NSP. O parâmetro S_{cd} indica o grau de ordenação/estruturação das cadeias alifáticas (caudas) dos lípidos de POPC na presença das NSP. Quanto mais próximo for de zero for o valor de S_{cd} , menor o alinhamento das cadeias alifáticas. É possível observar que as NSP induzem, de forma geral, um efeito estruturante ao longo das cadeias saturadas da bicamada (Figura 6.4, painel B), enquanto uma maior variabilidade é observada para as cadeias insaturadas (Figura 6.4, painel C). Nestas últimas, as curvas correspondentes às CS33_C e CS55_NI estão muito próximas da curva correspondente ao sistema de referência. A curva do composto CS12_C está ligeiramente acima e deslocada quando comparada com as restantes curvas na parte da cauda mais próxima das cabeça polar do lípido. Porém, o comportamento global das curvas é semelhante sugerindo um efeito tendencialmente estruturante das NSP.

O perfil do deslocamento quadrático médio (MSD) pode ser observado na Figura 6.4, painel A, e permite avaliar a difusão lateral das cadeias de POPC (sistema de referência) bem como o efeito provocado pela presença das moléculas de CS12_C, CS33_C e CS55_NI nesta propriedade. As curvas de MSD mostram que as 3 moléculas de NSP promovem uma diminuição da difusão lateral das cadeias de POPC quando comparadas com o sistema de referência (curva a preto no gráfico). Mais uma vez, e apesar de não ser notória uma diferença significativa de comportamento entre as moléculas de NSP, a curva referente ao sistema contendo a molécula de CS12_C (curva a azul) é a que induz uma maior alteração relativamente ao sistema de referência, sendo por isso, e tal como indicava o perfil de (S_{cd}) a

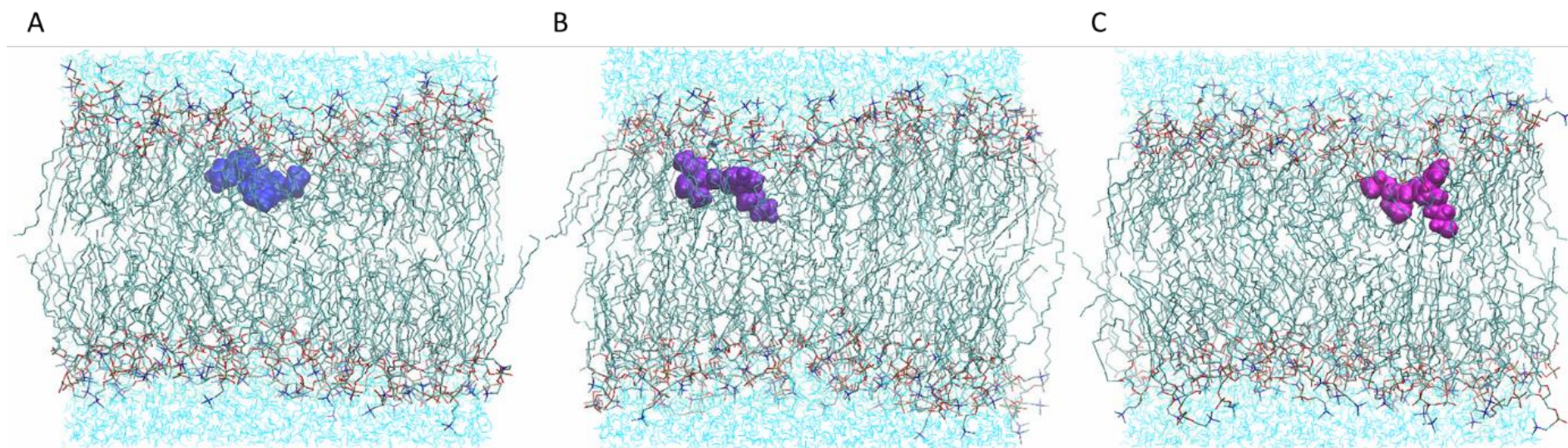


Figura 6.3 Configurações representativas do posicionamento das moléculas de NSP estudadas na bicamada de POPC. (A) CS12_C. (B) CS33_C. (C) CS55_NI

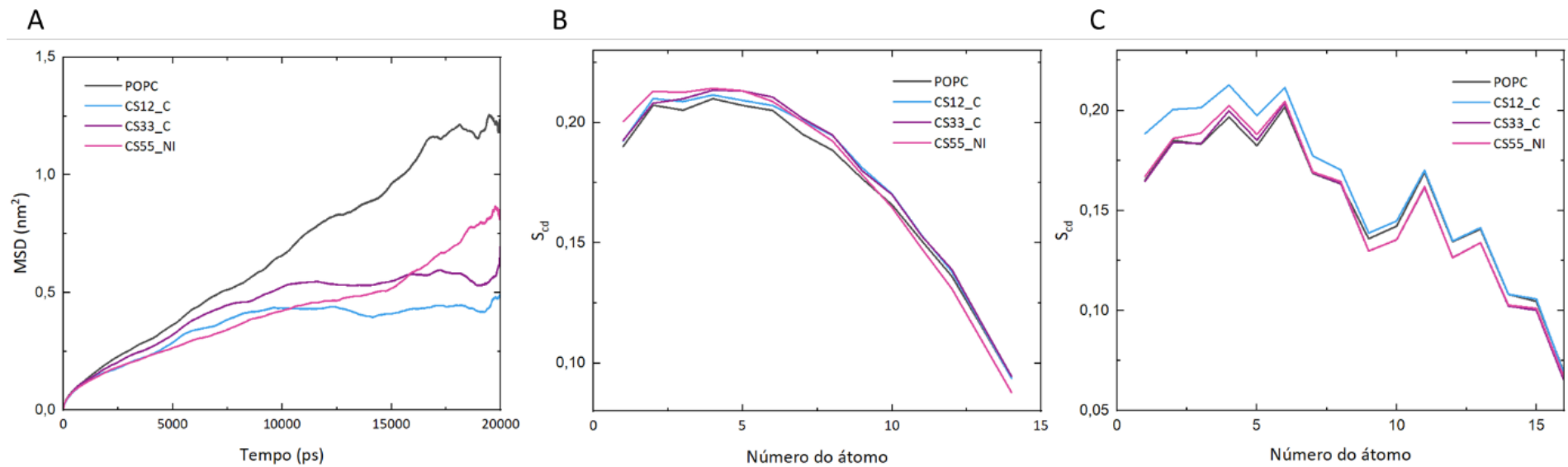


Figura 6.4 Resultados da análise de DM: (A) perfil de deslocamento quadrático médio (MSD) da membrana POPC na ausência e na presença das moléculas de NSP; (B) Parâmetro de ordem de deutério para a cauda saturada (S_{cd}), estimada ao longo da bicamada na ausência e presença das NSP; (C) Parâmetro de ordem de deutério para a cauda insaturada (S_{cd}), estimada ao longo da bicamada na ausência e na presença das NSP.

que promove um efeito de maior ordenação nas moléculas de lípido da bicamada. De salientar que as diferenças observadas entre sistemas, apontam para uma maior proximidade de comportamento entre as moléculas CS33_C e CS55_NI, tal como indicado pelo HCA, e uma maior diferença entre estas e a CS12_C, diferença essa que poderá eventualmente tornar-se mais evidente considerando futuramente sistemas mais concentrados (com maior número de moléculas de NSP em cada sistema).

6.2 Interação entre as moléculas de NSP e o recetor CB₁

O estudo da interação entre as três NSP selecionadas, CS12_C, CS33_C e CS55_NI, e o recetor CB₁ (PDB ID: 5XRA) foi realizada através do método *Docking* molecular. Nesta secção, encontram-se os resultados das conformações de menor energia obtidas nas simulações, cujos valores de energia de interação se encontram compilados na Tabela 6.1.

Tabela 6.1 Valores de energia de interação, E_{int} (Kcal/mol), correspondentes às configurações dos complexos entre as NSP selecionadas e o recetor CB₁ obtidos em cada simulação. dp corresponde ao desvio padrão.

	E_{int} (kcal.mol ⁻¹)			
	Simulação 1	Simulação 2	Simulação 3	$E_{int} \pm dp$
CS12_C	-12,54	-12,58	-12,59	-12,59 ± 0,04
CS33_C	-11,57	-11,42	-11,52	-11,52 ± 0,07
CS55_NI	-11,76	-12,26	-12,23	-12,1 ± 0,3

De acordo com a Tabela 6.1, os valores de energia de ligação entre simulações do mesmo sistema são muito próximos entre si e a estabilidade dos complexos segue a ordem CB₁: CS33_C < CB₁: CS55_NI < CB₁: CS12_C.

Na Tabela 6.2 encontram-se sumariados os resíduos do recetor CB₁ que estiveram preferencialmente envolvidos na interação com cada uma das moléculas NSP, bem como o tipo de interações dominantes que contribuem para a estabilização dos complexos.

Tabela 6.2 Lista de resíduos e interações responsáveis pela estabilização dos complexos entre o recetor CB₁ e as NSP, identificados por *docking* molecular. As interações identificadas incluem interações hidrofóbicas, ligações de hidrogénio convencionais e não convencionais. Os resíduos marcados a negrito correspondem aos identificados nas três simulações realizadas para cada molécula. Os resíduos assinalados a cor rosa correspondem a resíduos que são comuns aos Canabinoides CS33_C e CS55_NI e que não surgem no Canabinoide CS12_C para uma determinada interação. Por outro lado, os resíduos de cor azul referem-se a resíduos que interagem apenas com o compostos CS12_C.

CS12_C		CS33_C		CS55_NI	
Interações hidrofóbicas					
Simulação 1					
Phe170	Phe177	Phe170	Phe174	Val110	Val196
Val196	Phe200	Val196	Pro269	Ile267	Phe268
Ile267	Pro269	Phe379	Ala380	Pro269	Leu359
Trp356	Ala380			Phe379	Ala380
Cys386					
Simulação 2					
Phe170	Val196	Phe170	Phe174	Phe170	Phe174
Ile267	Pro269	His178	Phe200	His178	Phe268
Leu359	Ala380	Phe268	Trp279	Trp279	Leu359
Cys386		Lys376	Phe379	Lys376	Phe379
		Cys386		Ala380	Cys386
Simulação 3					
Phe170	Phe177	Phe170	Phe174	Phe174	Phe177
Val196	Phe200	His178	Val196	Ile267	Pro269
Pro269	Ala380	Phe200	Phe268	Trp279	Leu359
		Lys376	Phe379	Phe379	Ala380
		Ala380	Cys386		
Ligações de hidrogénio convencionais					
Simulação 1					
Ser383		His178		Ser383	Lys376
Simulação 2					
Ser383		Ser383	Ile267	Ser383	Phe108
				Ile267	Asp166
Simulação 3					
Ser383		Ser383	Asp266	His178	Ser383

		Ile267		Val196	
Interações de van der Waals					
Simulação 1					
Phe108	Ser173	Phe108	Ser173	Phe108	Phe170
Phe174	Phe189	Phe177	Phe189	Phe174	Phe189
Leu193	Thr197	Lys192	Leu193	His178	Leu193
Trp279	Phe268	Ile267	Phe268	Thr197	Phe200
Leu359	Met363	Ser383		Trp279	Met363
Lys376	Phe379			Thr377	Cys386
Simulação 2					
Phe108	Ser173			Val110	Phe177
Phe174	Phe189			Phe189	Leu193
Leu193	Phe268			Val196	Thr197
Met363	Lys376			Phe200	Pro269
Phe379				Met363	
Simulação 3					
Phe108	Ser173	Phe108	Phe177	Phe108	Val110
Phe174	Phe189	Phe189	Leu193	Phe170	Phe189
Leu193	Thr197	Thr197	Leu359	Leu193	Thr197
Ile267	Phe268	Met363		Phe200	Phe268
Trp279	Lys376			Met363	Lys376
Phe379					
Ligações de hidrogénio não convencionais					
Simulação 1					
		Lys376		His178	
Simulação 2					
		Val196	Ile267		
Simulação 3					

Os dados da Tabela 6.2 permitem concluir que no caso do complexo com o Canabinoide CS12_C, as interações hidrofóbicas mais comuns envolvem os resíduos Phe170, Val196, Pro269 e Ala380 do CB₁; a molécula de NSP estabelece ligações de hidrogénio com o resíduo Ser383 do recetor; e ainda as interações de *van der Waals* mais prevalentes foram estabelecidas com os resíduos Phe108, Ser173, Phe174, Phe189, Leu193, Lys376 e Phe379 do recetor.

No que diz respeito à molécula CS33_C, foi possível identificar essencialmente interações hidrofóbicas ocorrerem com os resíduos Phe170, Phe174 e Phe179 e as interações de *van der Waals* com os resíduos Phe108, Phe177, Phe189 e Leu193 do recetor.

No caso do Canabinoide CS55_NI as principais interações com o recetor CB₁ incluem os resíduos Leu359, Phe379 e Ala380 para interações hidrofóbicas, ligações de hidrogénio com o resíduo Ser383 e interações de *van der Waals* com os resíduos Phe189, Leu193, Thr197, Phe200 e Met363.

Um dos aspetos que importa salientar é a existência de interações com o recetor que são comuns entre as moléculas CS33_C e CS55_NI e que não foram identificadas no sistema contendo a molécula CS12_C (marcadas a rosa na tabela 6.2). Os resíduos do recetor CB₁ que estabelecem interações hidrofóbicas com as moléculas CS33_C e CS55_NI são o Phe174, His178, Phe268, Trp279, Lys376 e Phe379. Adicionalmente, o resíduo His178 pode estabelecer, também, uma ligação de hidrogénio com os dois canabinoides CS33_C e CS55_NI, o que não se verifica com o canabinoide CS12_C. Foi ainda possível identificar a existência de ligações de hidrogénio envolvendo o átomo de fluor dos canabinoides CS33_C e CS55_NI com os resíduos Ile267 e Asp266. Adicionalmente, verificou-se ainda o estabelecimento de ligações de hidrogénio não convencionais entre estas duas moléculas e o recetor CB₁, o que não acontece com o canabinoide CS12_C.

Nas Figuras 6.5 a 6.8 apresenta-se uma compilação dos resultados obtidos pelo *docking* molecular ilustrando as conformações mais favoráveis dos complexos, os modos de interação preferenciais, e os resíduos do sítio ativo do recetor CB₁ com maior afinidade para cada uma das moléculas de canabinoides estudadas.

Nestas figuras a representação no painel A refere-se à *Poseview*, ou seja, reflete a conformação bidimensional do sítio ativo do recetor com indicação com as interações mais prevalentes. Nos painéis B e C encontra-se, respetivamente, a representação bi e tridimensional com destaque de todos os resíduos envolvidos em cada tipo de interação.

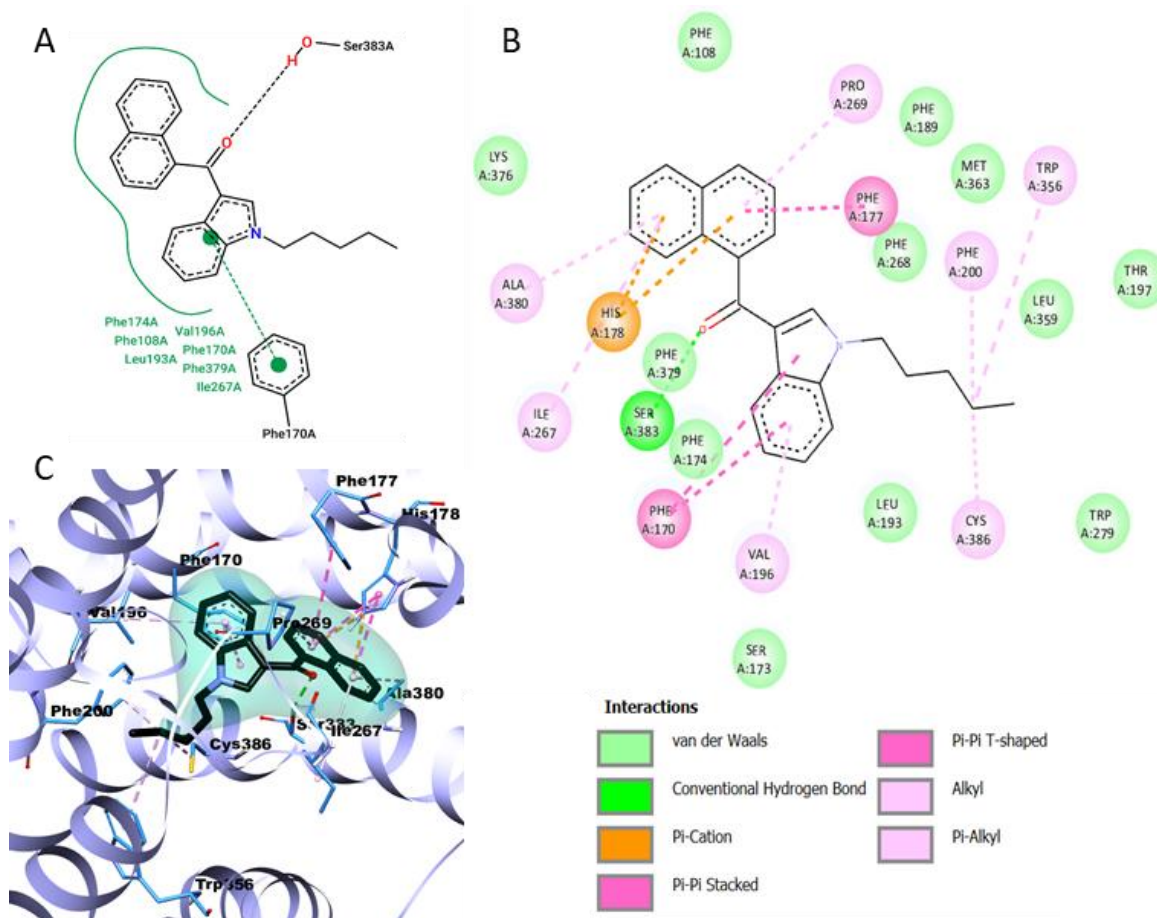


Figura 6.5 Conformação representativa das interações mais prevalentes estabelecidas entre o recetor CB₁ e o canabinoide CS12_C resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS12_C mostrando a ligação de hidrogénio estabelecida com o resíduo Ser383 e a ligação π - π com o resíduo Phe170, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional no qual estão sumariadas os tipos de interação mais prováveis entre o CS12_C e os resíduos do recetor CB₁. Resíduos rosa – interações hidrofóbicas, laranja - interações eletrostáticas, verde-claro - van der Waals, verde - ligações de hidrogénio convencionais; (C) Conformação mais estável do complexo (*Dock pose*) entre CS12_C (estrutura a preto) e do recetor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*.

O composto CS33_C apresentou dois modos de interação com o recetor, que diferem entre si na conformação da cadeia alifática que contém o átomo de flúor. O primeiro modo de interação encontra-se representado na Figura 6.6. No painel A é evidente a existência de uma ligação de hidrogénio entre o resíduo His178 do CB₁ e o grupo carbonilo do canabinoide mais próximo do grupo éter.

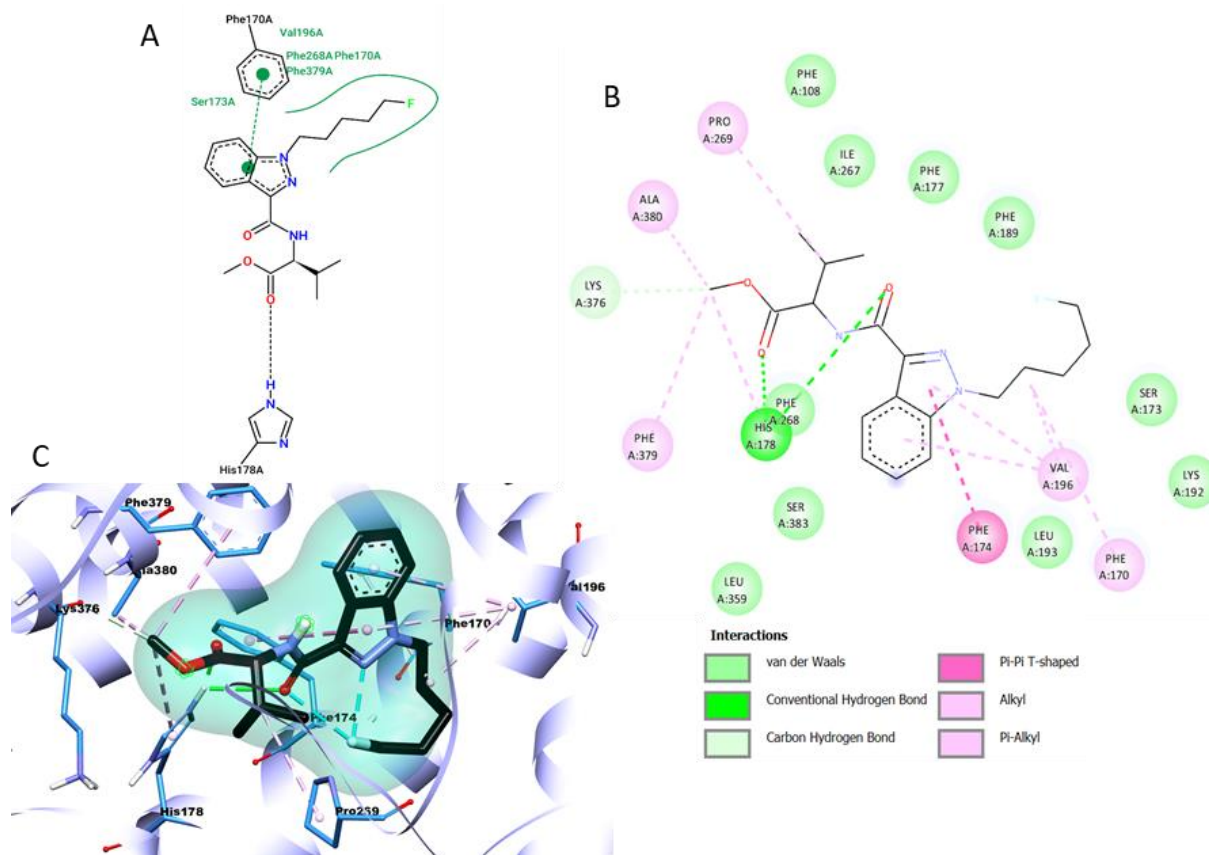


Figura 6.6 Conformação representativa de uma das interações mais favorável entre o recetor CB₁ e o canabinoide CS33_C resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS33_C mostrando a ligação de hidrogénio estabelecida com o resíduo His178 e a ligação π - π com o resíduo Phe170, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional no qual estão sumariadas os tipos de interação mais prováveis entre o CS33_C e os resíduos do recetor CB₁. Resíduos rosa – interações hidrofóbicas, verde-claro - van der Waals, verde - ligações de hidrogénio convencionais; (C) Conformação do complexo (*Dock pose*) entre CS33_C (estrutura a preto) e do recetor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*.

O segundo modo de interação observado entre o composto CS33_C e o recetor CB₁ encontra-se representado na Figura 6.7. No painel A, é possível identificar uma ligação de hidrogénio bifurcada entre o grupo OH do resíduo Ser383 e o átomo de oxigénio de um dos grupos carbonilo e o átomo de azoto do núcleo do canabinoide. No esquema representado painel B são visíveis em maior detalhe as interações e os resíduos envolvidos. É ainda notória a diferença da relativamente à conformação anterior devido à ausência da dobragem da cadeia alifática contendo o átomo de flúor identificada na Figura 6.6.

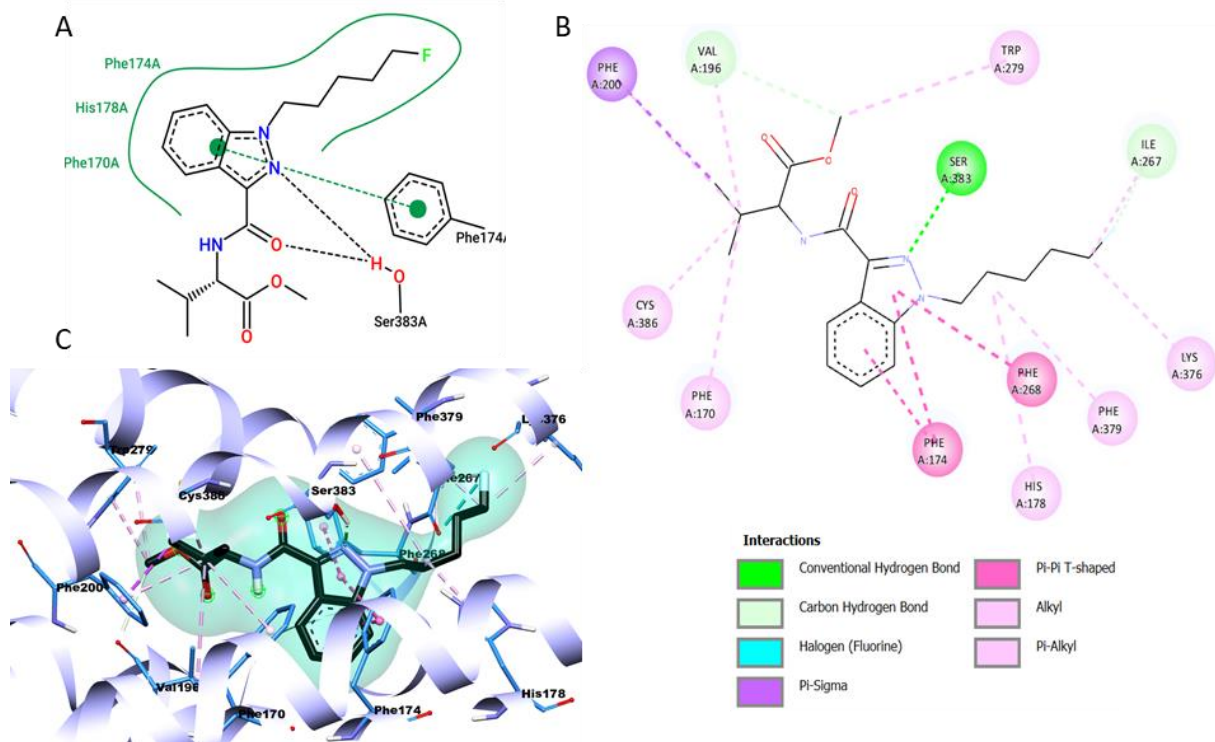


Figura 6.7 Conformação representativa de uma das interações mais favorável entre o recetor CB1 e o canabinoide CS33_C resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do recetor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS33_C mostrando a ligação de hidrogénio estabelecida com o resíduo Ser383 e a ligação π - π com o resíduo Phe174, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional sumariando os tipos de interação mais prováveis entre o CS33_C e os resíduos do recetor CB1. Resíduos rosa – interações hidrofóbicas, verde-claro – ligações de hidrogénio não convencionais, verde - ligações de hidrogénio convencionais, roxo – interações π - σ , azul – ligação com o halogénio; (C) Conformação do complexo (*Dock pose*) entre CS33_C (estrutura a preto) e do recetor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do recetor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*.

Na Figura 6.8 encontram-se sumariados os resultados obtidos para o canabinoide CS55_NI. No painel A estão identificadas duas ligações de hidrogénio, uma do tipo $NH \cdots O$ estabelecida entre o átomo de oxigénio do carbonilo do canabinoide e o resíduo His178 e outra do tipo $OH \cdots N$ estabelecida entre o resíduo Ser383 e o átomo de azoto do sistema aromático do canabinoide. Adicionalmente é possível identificar uma ligação do tipo $\pi \cdots \pi$ estabelecida entre o grupo aromático do canabinoide e o anel aromático do resíduo Phe170A e, tal como anteriormente, maiores detalhes são apresentados nos painéis B e C desta figura. No painel B é ainda sugerida a existência de uma ligação de hidrogénio adicional entre o átomo de oxigénio do grupo éter do canabinoide e o resíduo His178.

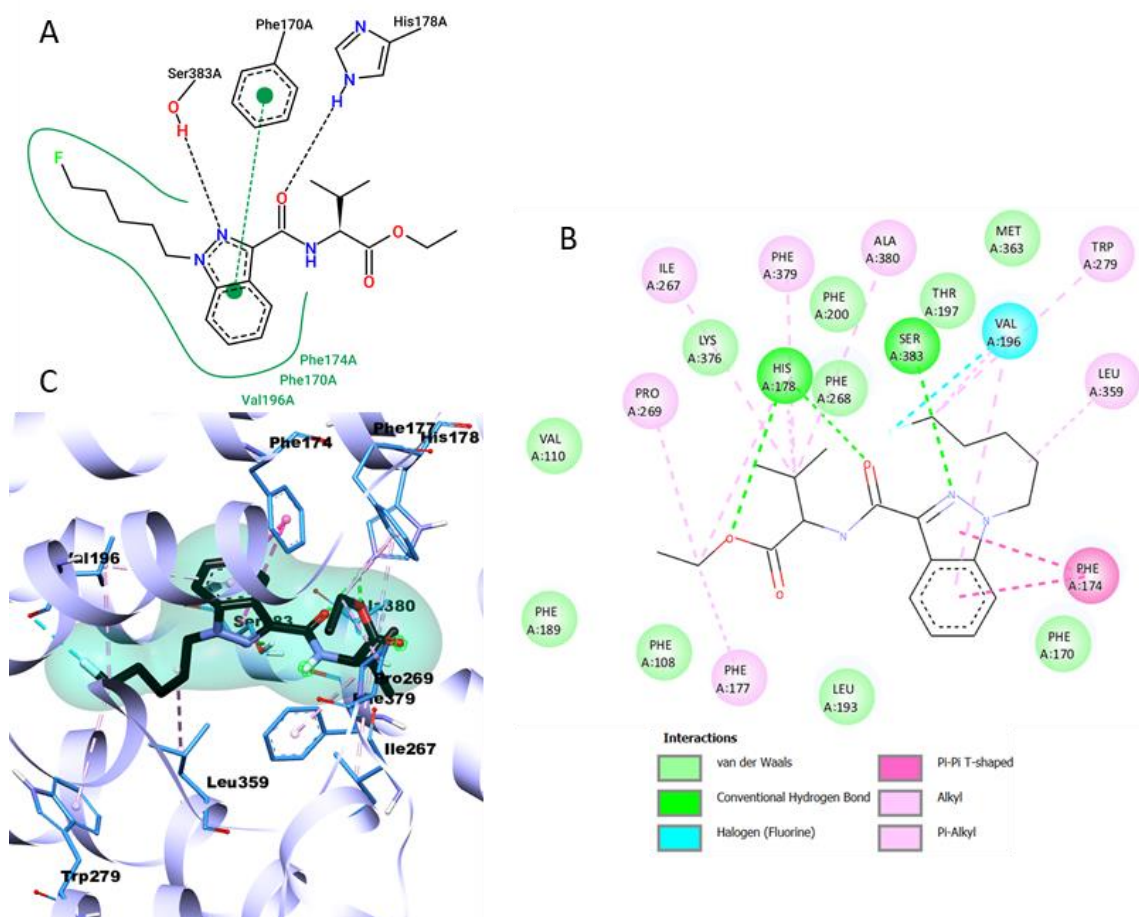


Figura 6.8 Conformação representativa das interações mais favoráveis entre o receptor CB1 e o canabinoide CS55_NI resultante da simulação por *docking* molecular. Em destaque encontram-se representados os resíduos do receptor mais relevantes localizados no seu sítio ativo. (A) *Poseview* do ligando CS55_NI mostrando as ligações de hidrogénio estabelecidas com os resíduos His178 e Ser383 e a ligação π - π com o resíduo Phe170, obtida com recurso ao servidor *Protein Plus* (<https://proteins.plus/>); (B) Diagrama bidimensional no qual estão sumariadas os tipos de interação mais prováveis entre o CS55_NI e os resíduos do receptor CB1. Resíduos rosa – interações hidrofóbicas, verde-claro – van der Waals, verde - ligações de hidrogénio convencionais, azul – ligação com o halogénio; (C) Conformação do complexo (*Dock pose*) entre CS55_NI (estrutura a preto) e do receptor (roxo) correspondente à classificação (*score*) mais elevada, ilustrando o modo de interação mais favorável entre os resíduos (azul) do receptor e o canabinoide. (B) e (C) foram obtidos recorrendo ao *software Discovery Studio Visualizer*.

De notar que as moléculas CS33_C e CS55_NI apresentam interações semelhantes entre si e mais distintas da molécula CS12_C. O canabinoide CS12_C estabelece apenas um tipo de ligação de hidrogénio com o resíduo Ser383 através do seu grupo carbonilo. No caso das moléculas CS33_C e CS55_NI, o mesmo tipo de interação pode ocorrer com os resíduos His178 e Ser383. De notar que o grupo carbonilo envolvido na ligação de hidrogénio com o resíduo Ser383 é comum nas duas moléculas CS33_C e CS55_NI, enquanto na molécula CS12_C o grupo carbonilo envolvido é o grupo próximo do grupo éter. As diferenças estruturais entre as moléculas de CS12_C e as moléculas CS33_C e CS55_NI verificam-se tanto na estrutura do núcleo aromático central como nos seus substituintes. O núcleo central é um grupo indole no caso de CS12_C e um grupo indazole no caso das moléculas CS33_C e CS55_NI.

Adicionalmente CS12_C possui uma cadeia lateral carbonada, enquanto nas moléculas CS33_C e CS55_NI esta cadeia contém um substituinte halogéneo. Por outro lado, o canabinoide CS12_C contém um grupo cetona e um grupo naftaleno enquanto os canabinoides CS33_C e CS55_NI, possuem um grupo cetona, um grupo amina e um grupo éster. De notar que a estrutura da molécula CS12_C apresenta menor liberdade conformacional, atendendo ao menor número de ângulos de torção existentes, o que lhe confere menor flexibilidade e maior restrição na interação com o recetor CB₁.

A natureza e força das interações responsáveis pela estabilização dos complexos entre os canabinoides e o recetor CB₁, identificadas nas simulações por *docking* molecular, foram ainda quantificadas pelo método do gradiente independente (IGM), tal como representado na Figura 6.9, na qual se discriminam as interações estabilizantes e destabilizantes.

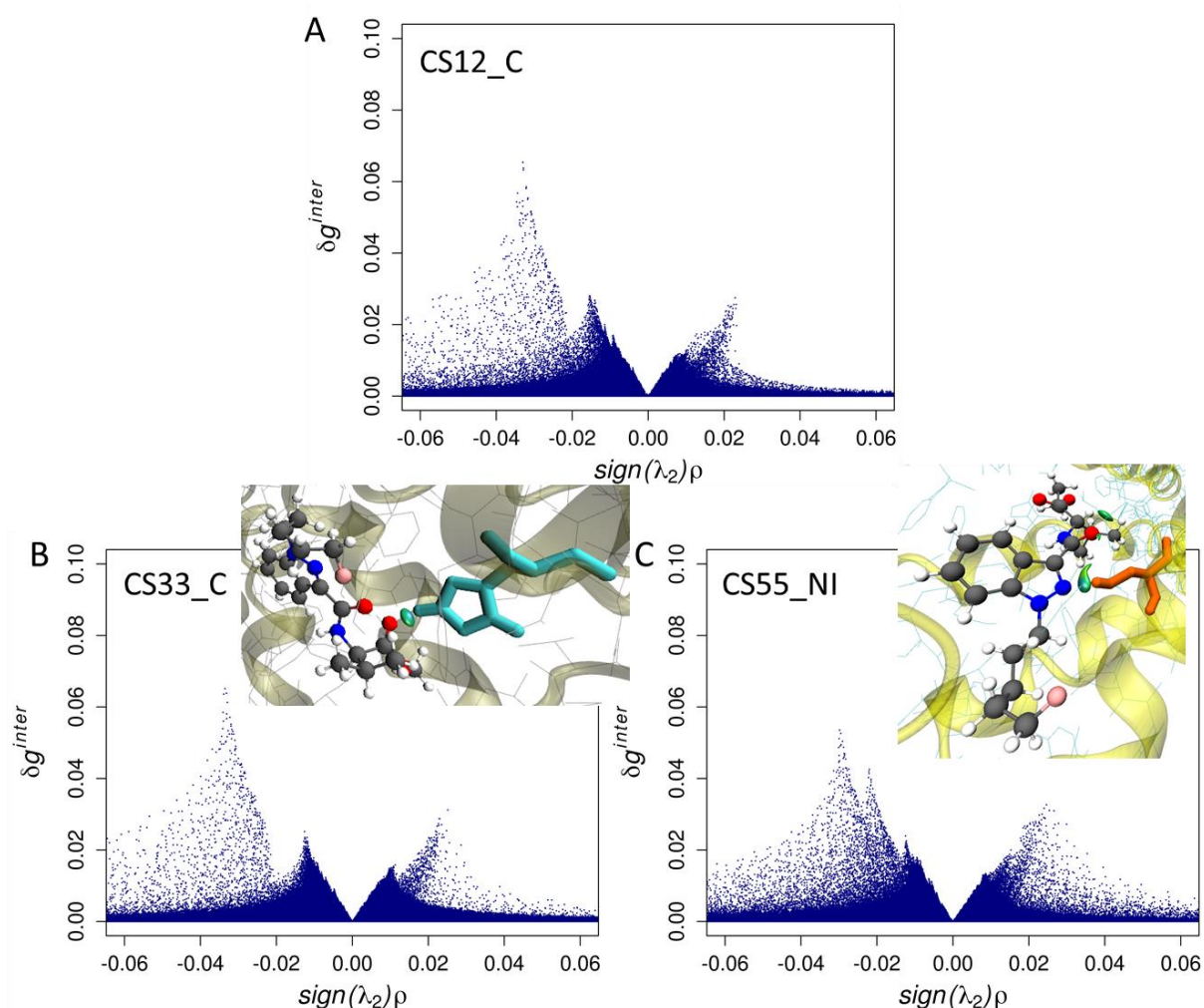


Figura 6.9 Visão geral dos padrões de interação dos complexos formados entre os canabinoides e o recetor CB₁, considerando a natureza e a força das interações intermoleculares identificadas nas conformações de menor energia obtidas nas simulações por *docking* molecular. (A) CS12_C:CB₁, (B) CS33_C:CB₁ e (C) CS55_NI:CB₁. As intensidades dos picos observadas nos gráficos 2D estão relacionadas com as energias de estabilização do complexo. Os contactos mais prevalentes entre os canabinoides e o recetor são também ilustrados pelas zonas de isosuperfície em 3D (para os compostos CS33_C e CS55_NI). Os sistemas presentes nas imagens dos esquemas (B) e (C) mostram as ligações de hidrogénio entre os compostos CS33_C e

CS55_NI com os resíduos His178 e Ser383, respetivamente. O esqueleto molecular dos compostos CS33_C e CS55_NI são representados de acordo com os seus átomos carbono, hidrogénio, oxigénio, azoto e fluor, pelas cores preto, branco, vermelho, azul e rosa, respetivamente. Os resíduos His178 e Ser383 estão representados pelas cores azul e laranja, respetivamente.

Para que a interação entre o recetor e cada uma das moléculas seja favorável, é necessário que as interações estabilizadoras sejam mais predominantes do que as destabilizantes. Este aspeto é patente na assimetria significativa entre os picos, localizados no lado negativo do eixo dos xx , correspondente às interações estabilizantes, e aos picos no lado direito, referentes às interações destabilizantes. Esta assimetria reflete um balanço favorável à formação dos complexos. Por outras palavras, os picos do lado negativo do eixo x ($sign(\lambda_2)\rho$), isto é, do lado estabilizador, possuem valores mais elevados de δg^{inter} , quando comparados com o lado oposto, indicando que o equilíbrio das interações não covalentes em cada sistema vão no sentido da formação de um complexo estável. Na representação gráfica 2D referente ao sistema CS12_C, painel A, o pico mais intenso ($\delta g^{inter} \approx 0.07$) diz respeito à interação mais provável, que corresponde a uma ligação de hidrogénio do tipo $O \cdots H - O$. De notar que os picos mais próximos da origem do eixo x , ($sign(\lambda_2)\rho$), correspondem a interações do tipo van der Waals. No caso da molécula CS33_C, painel B, o pico mais intenso corresponde à ligação de hidrogénio do tipo $O \cdots H - N$ estabelecida com o resíduo His178. No caso da molécula CS55_NI representada no painel C o pico mais intenso, à semelhança dos casos anteriores, corresponde a uma ligação de hidrogénio do tipo $N \cdots H - O$ estabelecida entre o composto CS55_NI e o resíduo Ser383. Adicionalmente destaca-se um segundo pico referente a uma ligação de hidrogénio do tipo $C - H \cdots N$ estabelecida com o resíduo His178.

Resumindo, os resultados obtidos quer da interação entre os canabinoides com uma membrana modelo, quer com o recetor CB₁ sugerem a existência de padrões de interação semelhantes entre as moléculas CS33_C e CS55_NI distinguindo-se do comportamento dos sistemas contendo a molécula CS12_C, corroborando os perfis estruturais identificados com os métodos quimiométricos.

Capítulo 7

Conclusão

7. Conclusão

A metodologia proposta neste estudo envolve a aprendizagem computacional e a simulação molecular como uma alternativa eficiente aos estudos convencionais para a caracterização de novas moléculas. Especificamente, foram explorados e discutidos detalhadamente dados relativos a 9 famílias de Novas Substâncias Psicoativas contendo moléculas ilícitas e outras ainda não regulamentadas com o objetivo de desenvolver novas ferramentas para caracterizar NSP desconhecidas tendo por base o seu perfil químico e a comparação com NSP já regulamentadas. A metodologia proposta tirou partido das potencialidades dos métodos não supervisionados de aprendizagem computacional e de métodos de simulação computacional atomística.

Os resultados obtidos mostraram a versatilidade desta combinação de métodos na classificação das NSP considerando o seu perfil químico e padrões de interação molecular. Foi possível (i) discriminar famílias de NSP e identificar moléculas pertencentes à mesma família, mas com comportamentos distintos, recorrendo a técnicas não supervisionadas, HCA e PCA (ii) identificar descritores moleculares redundantes, (iii) isolar e identificar grupos de compostos com interações específicas relevantes no meio biológico.

A abordagem proposta mostrou um grau de confiabilidade comparável a abordagens mais complexas sendo especialmente útil para promover a classificação rápida e económica de novas drogas psicoativas ainda não regulamentadas, sendo por isso um recurso valioso para melhorar o processo de triagem de novas drogas.

Referências bibliográficas

Referências bibliográficas

1. Plaa, Gabriel L. Toxicology. In : *Encyclopedia of Toxicology*. 2005. p. 347–350.
2. Dinis-Oliveira, Ricardo Jorge, Magalhães, Teresa and Carvalho, Félix Dias. Introdução à Toxicologia Forense. In : *Toxicologia Forense*. 2015. p. 1–8.
3. Adatsi, F. K. Forensic Toxicology. In : *Encyclopedia of Toxicology* [online]. Third Edit. Elsevier, 2014. p. 647–652. ISBN 9780123864543. <http://dx.doi.org/10.1016/B978-0-12-386454-3.00387-0>
4. Stout, Peter R. Forensic Toxicology. In : *Information Resources in Toxicology* [online]. Fourth Edi. Elsevier Inc., 2009. p. 293–296. ISBN 9780123735935. <http://dx.doi.org/10.1016/B978-0-12-373593-5.00032-X>
5. Wagner, Jarrad R. Introduction to forensic toxicology. In : *An Introduction to Interdisciplinary Toxicology*. 2020. p. 445–459. ISBN 9780128136027.
6. Mariotti, Kristiane de Cássia, Ortiz, Rafael Scorsatto and Limberger, Renata Pereira. Sequência Analítica em Toxicologia Forense. In : *Toxicologia Forense*. 2015. p. 109–130.
7. Tinoco, Rui. Para uma história dos efeitos das drogas: dos usos cerimoniais aos consumos malditos. *Antropológicas*. 2002. Vol. 6, p. 221–245.
8. Nunes, Laura M and Jólluskin, Gloria. O uso de drogas: breve análise histórica e social. *Revista da Faculdade de Ciências Humanas e Sociais* [online]. 2007. No. 4, p. 230–237. <https://bdigital.ufp.pt/bitstream/10284/449/1/230-237FCHS04-15.pdf>
9. Basu, Sutapa and Basu, Deeptanshu. The Relationship between Psychoactive Drugs, the Brain and Psychosis. *International Archives of Addiction Research and Medicine*. 2015. Vol. 1, p. 16–19. DOI 10.23937/2474-3631/1510003.
10. Utsumi, Natsuko. How Do Drugs and Hormones Influence Behavior? In : *Hormones, Brain and Behaviour*. 2001. p. 190–233.
11. OMS. *Neurociência do Uso e da Dependência de Substâncias Psicoativas*. 2004.
12. Sequeira, Margarida Maria Maniés. *Identificação e Quantificação de Novas Substâncias Psicoativas em Material Apreendido em Portugal*. 2018.
13. Bewley-Taylor, David and Jelsma, Martin. Regime change: Re-visiting the 1961 Single Convention on Narcotic Drugs. *International Journal of Drug Policy*. 2012. Vol. 23, p. 72–81. DOI 10.1016/j.drugpo.2011.08.003.
14. Zawilska, Jolanta B and Wojcieszak, Jakub. Novel Psychoactive Substances:

- Classification and General Information. In : *Synthetic Cathinones* [online]. Springer International Publishing, 2018. p. 11–24. ISBN 9783319787077. http://dx.doi.org/10.1007/978-3-319-78707-7_2
15. Torres, Ricardo Jorge Castelo de Sá. *Representações sociais das Novas Substâncias Psicoativas e da sua legislação*. 2015.
 16. Mercolini, Laura and Protti, Michele. Biosampling strategies for emerging drugs of abuse: towards the future of toxicological and forensic analysis. *Journal of Pharmaceutical and Biomedical Analysis* [online]. 2016. Vol. 130, p. 202–219. DOI 10.1016/j.jpba.2016.06.046.
 17. Silva, Joana Paula Pereira Barbosa Roque Da. *Novas Substâncias Psicoativas (NSP) nas festas e no ciberespaço: o retrato português de um fenómeno complexo*. 2018.
 18. Peacock, Amy, Bruno, Raimondo, Gisev, Natasa, Degenhardt, Louisa, Hall, Wayne, Sedefov, Roumen, White, Jason, Thomas, Kevin V and Farrell, Michael. New psychoactive substances: challenges for drug surveillance, control, and public health responses. *The Lancet* [online]. 2019. Vol. 394, p. 1668–1684. DOI 10.1016/S0140-6736(19)32231-7.
 19. EMCDDA. *Relatório Anual 2006: A Evolução do Fenómeno da Droga na Europa*. 2006.
 20. EMCDDA. *Relatório Anual 2007: A Evolução do Fenómeno da Droga na Europa 20*. 2007.
 21. EMCDDA. *Relatório Anual 2008: A evolução do fenómeno da droga na Europa* [online]. 2008. Available from: http://www.sicad.pt/BK/Publicacoes/Lists/SICAD_PUBLICACOES/Attachments/34/EMCDDA_AR2010_PT.pdf
 22. EMCDDA. *Relatório Anual 2009: A evolução do fenómeno da droga na Europa* [online]. 2009. Available from: http://www.sicad.pt/BK/Publicacoes/Lists/SICAD_PUBLICACOES/Attachments/34/EMCDDA_AR2010_PT.pdf
 23. EMCDDA. *Relatório anual 2010: A evolução do fenómeno da droga na Europa* [online]. 2010. Available from: http://www.sicad.pt/BK/Publicacoes/Lists/SICAD_PUBLICACOES/Attachments/34/EMCDDA_AR2010_PT.pdf
 24. EMCDDA. *Relatório anual 2011: A evolução do fenómeno da droga na Europa* [online]. 2011. Available from:

- http://www.sicad.pt/BK/Publicacoes/Lists/SICAD_PUBLICACOES/Attachments/34/EMCDDA_AR2010_PT.pdf
25. EMCDDA. *Relatório anual 2012: A evolução do fenómeno da droga na Europa* [online]. 2012. Available from: http://www.sicad.pt/BK/Publicacoes/Lists/SICAD_PUBLICACOES/Attachments/34/EMCDDA_AR2010_PT.pdf
 26. EMCDDA. *Relatório Europeu sobre Drogas 2013: Tendências e evoluções*. 2013.
 27. EMCDDA. *Relatório Europeu sobre Drogas 2014: Tendências e evoluções*. 2014.
 28. EMCDDA. *Relatório Europeu sobre Drogas 2015: Tendências e evoluções*. 2015.
 29. EMCDDA. *Relatório Europeu sobre Drogas 2016: Tendência e evolução* [online]. 2016. Available from: <https://www.emcdda.europa.eu/system/files/publications/2637/TDAT16001PTN.pdf>
 30. EMCDDA. *Relatório Europeu sobre Drogas 2017: Tendências e evoluções*. 2017.
 31. EMCDDA. *Relatório Europeu sobre Drogas 2018: Tendências e evoluções*. 2018.
 32. EMCDDA. *Relatório Europeu sobre Drogas 2019: Tendências e evoluções* [online]. 2019. Available from: http://www.emcdda.europa.eu/system/files/publications/11364/20191724_TDAT19001PTN_PDF.pdf
 33. EMCDDA. *European Drug Report 2020: Trends and Developments*. 2020.
 34. Madras, Bertha K. The Growing Problem of New Psychoactive Substances (NPS). *Current Topics Behavioral Neuroscience*. 2017. Vol. 32. DOI 10.1007/7854.
 35. Elliott, Simon and Evans, Julie. A 3-year review of new psychoactive substances in casework. *Forensic Science International* [online]. 2014. Vol. 243, p. 55–60. DOI 10.1016/j.forsciint.2014.04.017.
 36. Sainsbury, P. D., Kicman, A. T., Archer, R. P., King, L. A. and Braithwaite, R. A. Aminoindanes-the next wave of “legal highs”? *Drug Testing and Analysis*. 2011. Vol. 3, no. 7–8, p. 479–482. DOI 10.1002/dta.318.
 37. Shimshoni, Jakob A., Winkler, Ilan, Edery, Nir, Golan, Ezekiel, Van Wettum, René and Nutt, David. Toxicological evaluation of 5-methoxy-2-aminoindane (MEAI): Binge mitigating agent in development. *Toxicology and Applied Pharmacology*. 2017. Vol. 319, p. 59–68. DOI 10.1016/j.taap.2017.01.018.
 38. Manier, Sascha K., Felske, Christina, Eckstein, Niels and Meyer, Markus R. The metabolic

- fate of two new psychoactive substances – 2-aminoindane and N-methyl-2-aminoindane – studied in vitro and in vivo to support drug testing. *Drug Testing and Analysis*. 2020. Vol. 12, no. 1, p. 145–151. DOI 10.1002/dta.2699.
39. Pinterova, Nikola, Horsley, Rachel R. and Palenicek, Tomas. Synthetic aminoindanes: A summary of existing knowledge. *Frontiers in Psychiatry*. 2017. Vol. 8, no. NOV. DOI 10.3389/fpsyt.2017.00236.
 40. Alves, Vera L., Gonçalves, João L., Aguiar, Joselin, Teixeira, Helena M. and Câmara, José S. The synthetic cannabinoids phenomenon: from structure to toxicological properties. A review. *Critical Reviews in Toxicology* [online]. 2020. DOI 10.1080/10408444.2020.1762539.
 41. EMCDDA. Synthetic cannabinoids in Europe - update 2017. *Perspectives on Drugs* [online]. 2017. P. 9. Available from: <http://www.emcdda.europa.eu/topics/pods/synthetic-cannabinoids>
 42. Debruyne, Daniele and Le Boisselier, Reynald. Emerging drugs of abuse: current perspectives on synthetic cannabinoids. *Substance Abuse and Rehabilitation*. 2015. Vol. 6, p. 113–129. DOI 10.2147/sar.s73586.
 43. Seely, Kathryn A., Lapoint, Jeff, Moran, Jeffery H. and Fattore, Liana. Spice drugs are more than harmless herbal blends: A review of the pharmacology and toxicology of synthetic cannabinoids. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* [online]. 2012. Vol. 39, no. 2, p. 234–243. DOI 10.1016/j.pnpbp.2012.04.017.
 44. Couto, Rosa A.S., Gonçalves, Luís Moreira, Carvalho, Félix, Rodrigues, José A., Rodrigues, Cecília M.P. and Quinaz, M. Beatriz. The Analytical Challenge in the Determination of Cathinones, Key-Players in the Worldwide Phenomenon of Novel Psychoactive Substances. *Critical Reviews in Analytical Chemistry*. 2018. Vol. 48, no. 5, p. 372–390. DOI 10.1080/10408347.2018.1439724.
 45. Pieprzyca, Ewelina, Skowronek, Rafał, Nižnanský, Ľuboš and Czekaj, Piotr. Synthetic cathinones – From natural plant stimulant to new drug of abuse. *European Journal of Pharmacology*. 2020. Vol. 875. DOI 10.1016/j.ejphar.2020.173012.
 46. King, L. A. New phenethylamines in Europe. *Drug Testing and Analysis*. 2014. Vol. 6, p. 808–818. DOI 10.1002/dta.1570.
 47. Arsić, Biljana, Lukić, Vera, Micić, Ružica, Radosavljević, Živana and Nedović, Bojan.

- Overview of the major classes of new psychoactive substances, psychoactive effects, analytical determination and conformational analysis of selected illegal drugs. *Open Chemistry*. 2021. Vol. 19, p. 60–106. DOI 10.1515/chem-2021-0196.
48. Orsolini, Laura, Corkery, John M., Chiappini, Stefania, Guirguis, Amira, Vento, Alessandro, De Berardis, Domenico, Papanti, Duccio and Schifano, Fabrizio. 'New/Designer Benzodiazepines': An Analysis of the Literature and Psychonauts' Trip Reports. *Current Neuropharmacology*. 2020. Vol. 18, p. 809–837. DOI 10.2174/1570159x18666200110121333.
49. Zawilska, Jolanta B. and Wojcieszak, Jakub. An expanding world of new psychoactive substances—designer benzodiazepines. *NeuroToxicology* [online]. 2019. Vol. 73, p. 8–16. DOI 10.1016/j.neuro.2019.02.015.
50. Manchester, Kieran R., Lomas, Emma C., Waters, Laura, Dempsey, Fiona C. and Maskell, Peter D. The emergence of new psychoactive substance (NPS) Benzodiazepines: A review. *Drug Testing and Analysis*. 2018. Vol. 10, p. 37–53. DOI 10.1002/dta.2211.
51. Pérez-Mañá, Clara, Papaseit, Esther, Fonseca, Francina, Farré, Adriana, Torrens, Marta and Farré, Magi. Drug interactions with new synthetic opioids. *Frontiers in Pharmacology*. 2018. Vol. 9, no. OCT, p. 1–17. DOI 10.3389/fphar.2018.01145.
52. Beardsley, Patrick M. and Zhang, Yan. Synthetic opioids. *Handbook of Experimental Pharmacology*. 2018. P. 353–381. DOI 10.1007/164_2018_149.
53. Armenian, Patil, Vo, Kathy T., Barr-Walker, Jill and Lynch, Kara L. Fentanyl, fentanyl analogs and novel synthetic opioids: A comprehensive review. *Neuropharmacology* [online]. 2018. Vol. 134, p. 121–132. DOI 10.1016/j.neuropharm.2017.10.016.
54. Welz, Anna and Koba, Marcin. Piperazine derivatives as dangerous abused compounds. *Acta Pharmaceutica*. 2020. Vol. 70, no. 4, p. 423–441. DOI 10.2478/acph-2020-0035.
55. Wallach, Jason and Brandt, Simon D. Phencyclidine-Based New Psychoactive Substances. In : *Handbook of Experimental Pharmacology*. 2018.
56. Tittarelli, Roberta, Mannocchi, Giulio, Pantano, Flaminia and Romolo, Francesco. Recreational Use, Analysis and Toxicity of Tryptamines. *Current Neuropharmacology*. 2014. Vol. 13, no. 1, p. 26–46. DOI 10.2174/1570159x13666141210222409.
57. Araújo, Ana Margarida, Carvalho, Félix, Bastos, Maria de Lourdes, Guedes de Pinho, Paula and Carvalho, Márcia. The hallucinogenic world of tryptamines: an updated review. *Archives of Toxicology* [online]. 2015. Vol. 89, no. 8, p. 1151–1173.

- DOI 10.1007/s00204-015-1513-x.
58. Aryati, Widya Dwi, Winarko, Muhammad Siddiq, Susanto, Gerry May and Yanuar, Arry. Application of the machine and deep learning methods for the classification of cannabinoid- and cathinone-derived compounds. *International Journal of Applied Pharmaceutics*. 2020. Vol. 12, no. Special Issue 1, p. 47–50. DOI 10.22159/ijap.2020.v12s1.FF005.
 59. Rodrigues, Caio Henrique Pinke and Bruni, Aline Thaís. In silico toxicity as a tool for harm reduction: A study of new psychoactive amphetamines and cathinones in the context of criminal science. *Science and Justice*. 2018. Vol. 59, p. 234–247. DOI 10.1016/j.scijus.2018.11.006.
 60. Risoluti, R., Materazzi, S., Gregori, A. and Ripani, L. Early detection of emerging street drugs by near infrared spectroscopy and chemometrics. *Talanta* [online]. 2016. Vol. 153, p. 407–413. DOI 10.1016/j.talanta.2016.02.044.
 61. Omar, Jone, Slowikowski, Boleslaw, Guillou, Claude, Reniero, Fabiano, Holland, Margaret and Boix, Ana. Identification of new psychoactive substances (NPS) by Raman spectroscopy. *Journal of Raman Spectroscopy*. 2018. Vol. 50, p. 41–51. DOI 10.1002/jrs.5496.
 62. Metternich, Sonja, Fischmann, Svenja, Münster-Müller, Sascha, Pütz, Michael, Westphal, Folker, Schönberger, Torsten, Lyczkowski, Maren, Zörntlein, Siegfried and Huhn, Carolin. Discrimination of synthetic cannabinoids in herbal matrices and of cathinone derivatives by portable and laboratory-based Raman spectroscopy. *Forensic Chemistry* [online]. 2020. Vol. 19. DOI 10.1016/j.forc.2020.100241.
 63. Fowble, Kristen L., Shepard, Jason R.E. and Musah, Rabi A. Identification and classification of cathinone unknowns by statistical analysis processing of direct analysis in real time-high resolution mass spectrometry-derived “neutral loss” spectra. *Talanta* [online]. 2018. Vol. 179, p. 546–553. DOI 10.1016/j.talanta.2017.11.020.
 64. Gaspar, Helena, Bronze, Soraia, Oliveira, Catarina, Victor, Bruno L., Machuqueiro, Miguel, Pacheco, Rita, Caldeira, Maria João and Santos, Susana. Proactive response to tackle the threat of emerging drugs: Synthesis and toxicity evaluation of new cathinones. *Forensic Science International* [online]. 2018. Vol. 290, p. 146–156. DOI 10.1016/j.forsciint.2018.07.001.
 65. Ametovski, Adam, Macdonald, Christa, Manning, Jamie J., Haneef, S. A. Syed, Santiago,

- Marina, Martin, Lewis, Sparkes, Eric, Reckers, Andrew, Gerona, Roy R., Connor, Mark, Glass, Michelle and Banister, Samuel D. Exploring Stereochemical and Conformational Requirements at Cannabinoid Receptors for Synthetic Cannabinoids Related to SDB-006, 5F-SDB-006, CUMYL-PICA, and 5F-CUMYL-PICA. *ACS Chemical Neuroscience*. 2020. Vol. 11, no. 21, p. 3672–3682. DOI <https://doi.org/10.1021/acscchemneuro.0c00591>.
66. Vo, Quynh N., Mahinthichaichan, Paween, Shen, Jana and Ellis, Christopher R. How μ -opioid receptor recognizes fentanyl. *Nature Communications*. 2021. Vol. 12. DOI 10.1038/s41467-021-21262-9.
67. Olusegun, Atanda, Makun, Hussaini Anthony, Ogara, Isaac M., Edema, Mojisola, Idahor, Kingsley O., Oluwabamiwo, Bosede F. and Eshiett, Margaret E. Chemometrics: Theory and Application. In : *Multivariate Analysis in Management, Engineering and the Sciences* [online]. 2012. p. 121–132. Available from: <http://dx.doi.org/10.1039/C7RA00172J><https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics><http://dx.doi.org/10.1016/j.colsurfa.2011.12.014>
68. Kokot, Serge, Grigg, Michael, Panayiotou, Helen and Phuong, Tran Dong. Data Interpretation by some Common Chemometrics Methods. *Electroanalysis*. 1998. Vol. 10, p. 1081–1088. DOI 10.1002/(SICI)1521-4109(199811)10:16<1081::AID-ELAN1081>3.0.CO;2-X.
69. Hopke, Philip K. The evolution of chemometrics. *Analytica Chimica Acta*. 2003. Vol. 500, p. 365–377. DOI 10.1016/S0003-2670(03)00944-9.
70. Hibbert, David B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). In : *Pure and Applied Chemistry*. 2016. p. 407–443.
71. Héberger, Karoly. Chemoinformatics-multivariate mathematical-statistical methods for data evaluation. In : *Medical Applications of Mass Spectrometry*. 2008. p. 141–169. ISBN 9780444519801.
72. Pereira, L.S.A., Lisboa, Fernanda L.C., Coelho Neto, José, Valladão, Frederico N. and Sena, Marcelo M. Screening method for rapid classification of psychoactive substances in illicit tablets using mid infrared spectroscopy and PLS-DA. *Forensic Science International* [online]. 2018. Vol. 288, p. 227–235. DOI 10.1016/j.forsciint.2018.05.001.
73. Olesti, Eulàlia, De Toma, Ilario, Ramaekers, Johannes G., Brunt, Tibor M., Carbó, Marcel·lí, Fernández-Avilés, Cristina, Robledo, Patricia, Farré, Magí, Dierssen, Mara,

- Pozo, Óscar J. and De La Torre, Rafael. Metabolomics predicts the pharmacological profile of new psychoactive substances. *Journal of Psychopharmacology*. 2018. Vol. 33, p. 347–354. DOI 10.1177/0269881118812103.
74. Downs, GM and Barnard, JM. Clustering methods and their uses in computational chemistry. *Reviews in computational chemistry* [online]. 2002. Vol. 18, p. 1–40. Available from: http://books.google.com/books?hl=en&lr=&id=lqWXSz6QE8C&oi=fnd&pg=PA1&dq=Clustering+Methods+and+Their+Uses+in+Computational+Chemistry&ots=4_9lmOAcM9&sig=Fi-A1W8i9hnnEB7Unn4aYY9fDkE
75. Almeida, J. A.S., Barbosa, L. M.S., Pais, A. A.C.C. and Formosinho, S. J. Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemometrics and Intelligent Laboratory Systems*. 2007. Vol. 87, no. 2, p. 208–217. DOI 10.1016/j.chemolab.2007.01.005.
76. Kassambara, Alboukadel. Hierarchical Clustering. In : *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. 2017. p. 41–55.
77. Willett, Peter, Barnard, John M. and Downs, Geoffrey M. Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*. 1998. Vol. 38, no. 6, p. 983–996. DOI 10.1021/ci9800211.
78. Hierarchical Clustering/Dendrograms. In : *NCSS Statistical Software* [online]. 2019. p. 1–15. Available from: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Hierarchical_Clustering-Dendrograms.pdf
79. Alonso, Julio A. Principal Components Analysis: A Review of its Application on Molecular Dynamics Data. In : *ANNUAL REPORTS IN COMPUTATIONAL CHEMISTRY*. 2006. p. 233–261. ISBN 9781860947414.
80. Jolliffe, I. T. *Principal Component Analysis*. 2002.
81. Jolliffe, Ian T. and Cadima, Jorge. Principal component analysis: A review and recent developments. *Phil. Trans. R. Soc. A*. 2016. Vol. 374. DOI 10.1098/rsta.2015.0202.
82. Ganesan, Aravindhan, Coote, Michelle L. and Barakat, Khaled. Molecular dynamics-driven drug discovery: leaping forward with confidence. *Drug Discovery Today* [online]. 2017. Vol. 22, no. 2, p. 249–269. DOI 10.1016/j.drudis.2016.11.001.
83. Allen, Michael P. Introduction to Molecular Dynamics. *Computational Soft Matter:*

- From Synthetic Polymers to Proteins*. 2004. Vol. 23, p. 1–28.
84. Dar, Ayaz Mahmood and Mir, Shafia. Molecular Docking: Approaches, Types, Applications and Basic Challenges. *Journal of Analytical & Bioanalytical Techniques*. 2017. Vol. 08, no. 02, p. 8–10. DOI 10.4172/2155-9872.1000356.
85. Pagadala, Nataraj S., Syed, Khajamohiddin and Tuszynski, Jack. Software for molecular docking: a review. *Biophysical Reviews* [online]. 2017. Vol. 9, p. 91–102. DOI 10.1007/s12551-016-0247-1.
86. Hollingsworth, Scott A. and Dror, Ron O. Molecular Dynamics Simulation for All. *Neuron* [online]. 2018. Vol. 99, p. 1129–1143. DOI 10.1016/j.neuron.2018.08.011.
87. Justin A. Lemkul. GROMACS Tutorial. [online]. Available from: http://www.mdtutorials.com/gmx/membrane_protein/01_pdb2gmx.html
88. Leach, Andrew R. *Molecular modelling: Principles and Applications* [online]. 2001. ISBN 978-0-582-38210-7. Available from: <http://test1.bicpu.edu.in/binf/512c.pdf>
89. Van Gunsteren, Wilfred F., Bakowies, Dirk, Baron, Riccardo, Chandrasekhar, Indira, Christen, Markus, Daura, Xavier, Gee, Peter, Geerke, Daan P., Glättli, Alice, Hünenberger, Philippe H., Kastenholz, Mika A., Oostenbrink, Chris, Schenk, Merijn, Trzesniak, Daniel, Van Der Vegt, Nico F.A. and Yu, Haibo B. Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie - International Edition*. 2006. Vol. 45, p. 4064–4092. DOI 10.1002/anie.200502655.
90. Durrant, Jacob D and McCammon, J Andrew. Molecular dynamics simulations in drug discovery. *BMC Biology*. 2011. Vol. 9. DOI 10.1016/B978-0-12-809633-8.20154-4.
91. Ribeiro, Marcelino Da Costa. *Influência de Anestésico na Estrutura e Dinâmica de Membrana Celulares - Modelação e Simulação*. 2017.
92. Aittoniemi, Jussi, Niemelä, Perttu S., Hyvönen, Marja T., Karttunen, Mikko and Vattulainen, Ilpo. Insight into the putative specific interactions between cholesterol, sphingomyelin, and palmitoyl-oleoyl phosphatidylcholine. *Biophysical Journal*. 2007. Vol. 92, p. 1125–1137. DOI 10.1529/biophysj.106.088427.
93. Morris, Garret M. and Lim-Wilby, Marguerita. Molecular Docking. In : *Methods in Molecular Biology*. 2006. p. 365–382.
94. Roy, Kunal, Kar, Supratik and Das, Rudra Narayan. Other Related Techniques. In : *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. 2015. p. 357–425. ISBN 9780128015056.

95. Wang, Guimin and Zhu, Weiliang. Molecular docking for drug discovery and development: A widely used approach but far from perfect. *Future Medicinal Chemistry*. 2016. Vol. 8, no. 14. DOI 10.4155/fmc-2016-0143.
96. Lefebvre, Corentin, Rubez, Gaëtan, Khartabil, Hassan, Boisson, Jean Charles, Contreras-García, Julia and Hénon, Eric. Accurately extracting the signature of intermolecular interactions present in the NCI plot of the reduced density gradient versus electron density. *Physical Chemistry Chemical Physics*. 2017. Vol. 19, p. 17928–17936. DOI 10.1039/c7cp02110k.
97. Lefebvre, Corentin, Khartabil, Hassan, Boisson, Jean Charles, Contreras-García, Julia, Piquemal, Jean Philip and Hénon, Eric. The Independent Gradient Model: A New Approach for Probing Strong and Weak Interactions in Molecules from Wave Function Calculations. *ChemPhysChem*. 2018. Vol. 19, p. 724–735. DOI 10.1002/cphc.201701325.
98. Mendes, Maria, Cova, Tânia, Basso, João, Ramos, M. Luísa, Vitorino, Rui, Sousa, João, Pais, Alberto and Vitorino, Carla. Hierarchical design of hyaluronic acid-peptide constructs for glioblastoma targeting: Combining insights from NMR and molecular dynamics simulations. *Journal of Molecular Liquids* [online]. 2020. Vol. 315, p. 1–12. DOI 10.1016/j.molliq.2020.113774.
99. Cova, Tânia F., Milne, Bruce F. and Pais, Alberto A.C.C. Host flexibility and space filling in supramolecular complexation of cyclodextrins: A free-energy-oriented approach. *Carbohydrate Polymers* [online]. 2019. Vol. 205, p. 42–54. DOI 10.1016/j.carbpol.2018.10.009.
100. International Narcotics Control Board (INCB). *List of Psychotropic Substances under International Control* [online]. 2020. Available from: <https://www.incb.org/>
101. EMCDDA. *EMCDDA initial report on the new psychoactive substance methyl 3,3-dimethyl-2-(1-(pent-4-en-1-yl)-1H-indazole-3-carboxamido)butanoate (MDMB-4en-PINACA)*. 2020.
102. UNODC. List of Announcements. [online]. Available from: <https://www.unodc.org/LSS/Announcement?type=NPS>
103. Cova, Tânia F.G.G., Pais, Alberto A.C.C. and Formosinho, Sebastião J. Iberian universities: A characterisation from ESI rankings. *Scientometrics*. 2013. Vol. 94, p. 1239–1251. DOI 10.1007/s11192-012-0792-8.

104. Jagannathan, Ramesh. Identification of Psychoactive Metabolites from Cannabis sativa, Its Smoke, and Other Phytocannabinoids Using Machine Learning and Multivariate Methods. *ACS Omega*. 2020. Vol. 5, p. 281–295. DOI 10.1021/acsomega.9b02663.
105. Mendes, Maria, Basso, João, Silva, Jessica, Cova, Tânia, Sousa, João, Pais, Alberto and Vitorino, Carla. Biomimetic ultra-small lipid nanoconstructs for glioblastoma treatment: A computationally guided experimental approach. *International Journal of Pharmaceutics* [online]. 2020. DOI 10.1016/j.ijpharm.2020.119661.
106. University of Calgary. Biocomputing Group. *Using computer simulations to study biological problems*.
107. Malde, Alpeshkumar K., Zuo, Le, Breeze, Matthew, Stroet, Martin, Poger, David, Nair, Pramod C., Oostenbrink, Chris and Mark, Alan E. An Automated force field Topology Builder (ATB) and repository: Version 1.0. *Journal of Chemical Theory and Computation*. 2011. Vol. 7, p. 4026–4037. DOI 10.1021/ct200196m.
108. Berger, Oliver, Edholm, Olle and Jahnig, Fritz. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophysical Journal*. 1997. Vol. 72, p. 2002–2013. DOI 10.1016/S0006-3495(97)78845-3.
109. Pettus, Kelsey Leigh. *Study of an Active-State CB1 Receptor Model and JWH Compound Interactions to Predict New Emerging Synthetic Cannabinoids*. 2016.
110. Shao, Zhenhua, Yan, Wei, Chapman, Karen, Ramesh, Karthik, Ferrell, Aaron J., Yin, Jie, Wang, Xuehui, Xu, Qingping and Rosenbaum, Daniel M. Structure of an allosteric modulator bound to the CB1 cannabinoid receptor. *Nature Chemical Biology* [online]. 2019. Vol. 15, no. 12, p. 1199–1205. DOI 10.1038/s41589-019-0387-2.
111. Hua, Tian, Vemuri, Kiran, Nikas, Spyros P., Laprairie, Robert B., Wu, Yiran, Qu, Lu, Pu, Mengchen, Korde, Anisha, Jiang, Shan, Ho, Jo Hao, Han, Gye Won, Ding, Kang, Li, Xuanxuan, Liu, Haiguang, Hanson, Michael A., Zhao, Suwen, Bohn, Laura M., Makriyannis, Alexandros, Stevens, Raymond C. and Liu, Zhi Jie. Crystal structures of agonist-bound human cannabinoid receptor CB 1. *Nature*. 2017. Vol. 547, no. 7664, p. 468–471. DOI 10.1038/nature23272.
112. Jung, Sang Won, Cho, Art E. and Yu, Woogyung. Exploring the Ligand Efficacy of Cannabinoid Receptor 1 (CB1) using Molecular Dynamics Simulations. *Scientific Reports* [online]. 2018. Vol. 8, no. 1, p. 1–11. DOI 10.1038/s41598-018-31749-z.

113. Tkachenko, Volodymyr, Farafonov, Vladimir, Tokarev, Viktor and Tkachenko, Irina. Study of the effectiveness of various cannabinoid receptor 1 (CB1) agonists using molecular docking and molecular dynamics modeling. *French-Ukrainian Journal of Chemistry*. 2020. Vol. 8, no. 1, p. 76–87. DOI 10.17721/fujcv8i1p76-87.
114. Tham, Mylyne, Yilmaz, Orhan, Alaverdashvili, Mariam, Kelly, Melanie E.M., Denovan-Wright, Eileen M. and Laprairie, Robert B. Allosteric and orthosteric pharmacology of cannabidiol and cannabidiol-dimethylheptyl at the type 1 and type 2 cannabinoid receptors. *British Journal of Pharmacology*. 2019. Vol. 176, no. 10, p. 1455–1469. DOI 10.1111/bph.14440.
115. Tuccinardi, Tiziano, Ferrarini, Pier Luigi, Manera, Clementina, Ortore, Gabriella, Saccomanni, Giuseppe and Martinelli, Adriano. Cannabinoid CB2/CB1 selectivity. Receptor modeling and automated docking analysis. *Journal of Medicinal Chemistry*. 2006. Vol. 49, p. 984–994. DOI 10.1021/jm050875u.
116. Mella-Raipán, Jaime A., Lagos, Carlos F., Recabarren-Gajardo, Gonzalo, Espinosa-Bustos, Christian, Romero-Parra, Javier, Pessoa-Mahana, Hernán, Iturriaga-Vásquez, Patricio and Pessoa-Mahana, Carlos David. Design, synthesis, binding and docking-based 3D-QSAR studies of 2-pyridylbenzimidazoles - A new family of high affinity CB1 cannabinoid ligands. *Molecules*. 2013. Vol. 18, no. 4, p. 3972–4001. DOI 10.3390/molecules18043972.
117. Stein, Reed. *Understanding virtual solvent through large-scale ligand discovery* [online]. 2020. Available from: <https://escholarship.org/uc/item/4b7485k5>
118. Durdagi, Serdar, Papadopoulos, Manthos G., Zoumpoulakis, Panagiotis G., Koukoulitsa, Catherine and Mavromoustakos, Thomas. A computational study on cannabinoid receptors and potent bioactive cannabinoid ligands: Homology modeling, docking, de novo drug design and molecular dynamics analysis. *Molecular Diversity*. 2010. Vol. 14, p. 257–276. DOI 10.1007/s11030-009-9166-4.
119. Ai, Rizi and Chang, Chia En A. *Ligand-specific homology modeling of human cannabinoid (CB1) receptor*. 2012.
120. Hua, Tian, Vemuri, Kiran, Pu, Mengchen, Qu, Lu, Han, Gye Won, Wu, Yiran, Zhao, Suwen, Shui, Wenqing, Li, Shanshan, Korde, Anisha, Laprairie, Robert B., Stahl, Edward L., Ho, Jo Hao, Zvonok, Nikolai, Zhou, Han, Kufareva, Irina, Wu, Beili, Zhao, Qiang, Hanson, Michael A., Bohn, Laura M., Makriyannis, Alexandros, Stevens, Raymond C. and

- Liu, Zhi Jie. Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell*. 2016. Vol. 167, p. 750–762. DOI 10.1016/j.cell.2016.10.004.
121. Hua, Tian, Li, Xiaoting, Wu, Lijie, Iliopoulos-Tsoutsouvas, Christos, Wang, Yuxia, Wu, Meng, Shen, Ling, Brust, Christina A., Nikas, Spyros P., Song, Feng, Song, Xiyong, Yuan, Shuguang, Sun, Qianqian, Wu, Yiran, Jiang, Shan, Grim, Travis W., Benchama, Othman, Stahl, Edward L., Zvonok, Nikolai, Zhao, Suwen, Bohn, Laura M., Makriyannis, Alexandros and Liu, Zhi Jie. Activation and Signaling Mechanism Revealed by Cannabinoid Receptor-Gi Complex Structures. *Cell* [online]. 2020. Vol. 180, p. 655–665. DOI 10.1016/j.cell.2020.01.008.
122. Zagzoog, Ayat, Brandt, Asher L., Black, Tallan, Kim, Eunhyun D., Burkart, Riley, Patel, Mikin, Jin, Zhiyun, Nikolaeva, Maria and Laprairie, Robert B. Assessment of select synthetic cannabinoid receptor agonist bias and selectivity between the type 1 and type 2 cannabinoid receptor. *Scientific Reports* [online]. 2021. Vol. 11, p. 1–18. DOI 10.1038/s41598-021-90167-w.

Anexos

Anexo 1

- Descritores Híbridos

Tabela A 1 Descritores presentes na categoria dos Descritores Híbridos. Nesta categoria existem 2 classes de descritores.

WHIM			BCUT
Wlambda1.unity	Wlambda2.unity	Wlambda3.unity	BCUTw.1l
Wgamma1.unity	Wgamma2.unity	Wgamma3.unity	BCUTw.1h
Wnu1.unity	Wnu2.unity	Weta2.unity	BCUTc.1l
Weta3.unity	Weta1.unity	WT.unity	BCUTc.1h
WA.unity	WG.unity	WV.unity	BCUTp.1l
WD.unity	WK.unity		BCUTp.1h

- Descritores Constitucionais

Tabela A 2 Descritores presentes na categoria dos Descritores Constitucionais. Nesta categoria existem 16 classes de descritores.

FractionalCSP3	Fsp3	XLogP	XLogP
Weight	MW	RotatableBondsCount	nRotB
MannholdLogP	MLogP	RuleOfFive	LipinskiFailures
LongestAliphaticChain	nAtomLAC	LargestPiSystem	nAtomP
BondCount	nB	LargestChain	nAtomLC
BasicGroupCount	nBase	AtomCount	nAtom
ALOGP	AlogP	AcidicGroupCount	nAcid
	AlogP2	AromaticAtomsCount	naAromAtom
	AMR	AromaticBondsCount	nAromBond

- Descritores Eletrônicos

Tabela A 3 Descritores presentes na categoria dos Descritores Eletrônicos. Nesta categoria existem 5 classes de descritores.

HBondDonorCount			HBondAcceptorCount	
nHBDon			nHBAcc	
CPSA				
PPSA.1	PPSA.2	PPSA.3	PNSA.1	PNSA.2
PNSA.3	DPSA.1	DPSA.2	DPSA.3	FPSA.1
FPSA.2	FPSA.3	FNSA.1	FNSA.2	FNSA.3

WPSA.1	WPSA.2	WPSA.3	WNSA.1	WNSA.2
WNSA.3	RPCG	RNCG	RPCS	RNCS
THSA	TPSA	RHSA	RPSA	
BPol			APol	
bpol			apol	

- **Descritores Topológicos**

Tabela A 4 Descritores presentes na categoria dos Descritores Topológicos. Nesta categoria existem 25 classes de descritores.

AutocorrelationPolarizability		AutocorrelationMass		AutocorrelationCharge	
ATSp1		ATSm1		ATSc1	
ATSp2		ATSm2		ATSc2	
ATSp3		ATSm3		ATSc3	
ATSp4		ATSm4		ATSc4	
ATSp5		ATSm5		ATSc5	
ChiCluster		ChiChain		CarbonTypes	
SC.3	SC.4	SCH.3	SCH.4	C1SP1	C1SP2
SC.5	SC.6	SCH.5	SCH.6	C2SP1	C2SP2
VC.3	VC.4	SCH.7	VCH.3	C3SP2	C1SP3
VC.5	VC.6	VCH.4	VCH.5	C2SP3	C3SP3
ChiPathCluster		VCH.6	VCH.7	C4SP3	
SPC.4	SPC.5	ChiPath			
SPC.6	VPC.4	SP.0	SP.1	SP.2	SP.3
VPC.5	VPC.6	SP.4	SP.5	SP.6	SP.7
SmallRing		VP.0	VP.1	VP.2	VP.3
nSmallRings	nAromRings	VP.4	VP.5	VP.6	VP.7
nRingBlocks	nAromBlocks	MDED			
nRings3	nRings4	MDEC.11	MDEC.12	MDEC.13	MDEC.14
nRings5	nRings	MDEC.22	MDEC.23	MDEC.24	MDEC.33
nRings7	nRings8	MDEC.34	MDEC.44	MDEO.11	MDEO.12
nRings9		MDEO.22	MDEN.11	MDEN.12	MDEN.13
VAdjMat	VAdjMat	MDEN.22	MDEN.23	MDEN.33	
FractionalPSA	ZagrebIndex	WienerNumbers		VABC	TPSA

tpsaEfficiency	Zagreb	WPATH	WPOL	VABC	topoPSA
KappaShapeIndices			HybridizationRatio	FMF	
Kier1	Kier2	Kier3	HybRatio		FMF
PetitjeanNumber		FragmentComplexity		EccentricConnectivityIndex	
PetitjeanNumber		fragC		ECCEN	
WeightedPath				PetitjeanShapeIndex	
WTPT.1	WTPT.2	WTPT.3		topoShape	
WTPT.4	WTPT.5			geomShape	
KierHallSmarts					
khs.sLi	khs.ssBe	khs.ssssBe	khs.ssBH	khs.sssB	khs.ssssB
khs.sCH3	khs.dCH2	khs.ssCH2	khs.tCH	khs.dsCH	khs.aaCH
khs.sssCH	khs.ddC	khs.tsC	khs.dssC	khs.aasC	khs.aaa
khs.ssssC	khs.sNH3	khs.sNH2	khs.ssNH2	khs.dN	khs.ssNH
khs.aaNH	khs.tN	khs.sssNH	khs.dsN	khs.aaN	khs.sssN
khs.ddsN	khs.aasN	khs.ssssN	khs.sOH	khs.dO	khs.ssO
khs.aaO	khs.sF	khs.sSiH3	khs.ssSiH2	khs.sssSiH	khs.ssssSi
khs.sPH2	khs.ssPH	khs.sssP	khs.dsssP	khs.sssssP	khs.sSH
khs.dS	khs.ssS	khs.aaS	khs.dssS	khs.ddssS	khs.sCl
khs.sGeH3	khs.ssGeH2	khs.sssGeH	khs.ssssGe	khs.sAsH2	khs.ssAsH
khs.sssAs	khs.ssdAs	khs.ssssAs	khs.sSeH	khs.dSe	khs.ssSe
khs.aaSe	khs.dssSe	khs.ddssSe	khs.sBr	khs.sSnH3	khs.ssSnH2
khs.sssSnH	khs.ssssSn	khs.sI	khs.sPbH3	khs.ssPbH2	khs.sssPbH
khs.ssssPb					

- Descritores Geométricos

Tabela A 5 Descritores presentes na categoria dos Descritores Geométricos. Nesta categoria existem 3 classes de descritores.

MomentOfInertia			GravitationalIndex			LengthOverBreadth
MOMI.X	MOMI.Y	MOMI.Z	GRAV.1	GRAV.2	GRAV.3	LOBMAX
MOMI.XY	MOMI.XZ	MOMI.YZ	GRAVH.1	GRAVH.2	GRAVH.3	LOBMIN
MOMI.R			GRAV.4	GRAV.5	GRAV.6	

Anexo 2

Tabela A 6 Definição dos descritores pertencentes ao conjunto final.

Descritor	Definição
apol	Soma das polarizabilidades atómicas
ATSm1	Calcula a autocorrelação, onde o peso é igual à massa atómica normalizada
BCUT.1I	Tem por base a conectividade e as propriedades atómicas da molécula, devolvendo o valor próprio mais elevado para a carga parcial mais baixa
bpol	Soma do valor absoluto da diferença entre as polarizabilidades atómicas de todos os átomos ligados na molécula
C1SP2	Carbono com hibridização sp2 ligado a outro carbono
C1SP3	Carbono com hibridização sp3 ligado a outro carbono
C2SP2	Carbono com hibridização sp2 ligado a outros dois carbonos
C2SP3	Carbono com hibridização sp3 ligado a outros dois carbonos
C3SP2	Carbono com hibridização sp2 ligado a outros três carbonos
C3SP3	Carbono com hibridização sp3 ligado a outros três carbonos
C4SP3	Carbono com hibridização sp3 ligado a outros quatro carbonos
FMF	Caracteriza a complexidade da molécula
FPSA.3	Razão entre a área da superfície da carga parcial positiva e a superfície molecular total
khs.aaN	Representa o fragmento: [N,nD2H0](:*):*
khs.aasN	Representa o fragmento: [N,nD3H0](:*)(:*)-,:*
khs.dsN	Representa o fragmento: [ND2H0](=*)-*
khs.sCl	Representa o fragmento: [ClD1]-*
khs.sNH2	Representa o fragmento: [ND2H2](-*)-*
khs.ssNH	Representa o fragmento: [ND2H](-*)-*
MDEN.12	Distância entre todos os nitrogénios primários e secundários
MDEN.13	Distância entre todos os nitrogénios primários e terciários

MDEO.11	Distância entre todos os oxigénios primários
MDEO.12	Distância entre todos os oxigénios primários e secundários
MLogP	Previsão do valor de logP tendo por base no número de átomos de carbono e heteroátomos
MOMI.XY	Razão dos valores do momento de inércia entre os eixos X e Y
MW	Peso molecular
nAromAtom	Número de átomos aromáticos numa molécula
nAromBond	Número de ligações aromáticas numa molécula
nAtomLAC	Número de átomos da maior cadeia alifática
nAtomLC	Número de átomos da maior cadeia
nAtomP	Número de átomos no maior sistema π
nB	Número de ligações numa molécula
nBase	Número de grupos básicos, estes são representados pelos fragmentos $[\$(\text{NH}_2)-[\text{CX}_4]]$, $[\$(\text{NH})(-[\text{CX}_4])-\text{CX}_4]]$, $[\$(\text{N}(-[\text{CX}_4])(-[\text{CX}_4])-\text{CX}_4)]$, $[\$(\text{[*];+;!$(*~[*;-])]]$, $[\$(\text{N}=\text{C}-\text{N})]$ e $[\$(\text{N}-\text{C}=\text{N})]$
nHBAcc	Número de átomos aceitadores de ligações de hidrogénio
nHBDon	Número de átomos dadores de ligações de hidrogénio
nRingBlocks	Número total de blocos de anéis distintos
nRings3	Número de anéis com três átomos
nRings5	Número de anéis com cinco átomos
nRings6	Número de anéis com seis átomos
nRings7	Número de anéis com sete átomos
nSmallRings	Número de anéis com três a nove átomos
TopoPSA	Área topológica da superfície polar
VCH.7	Cadeia de valência de ordem 7

Anexo 3

Tabela A 7 Lista de resíduos do recetor que interagiu com os compostos em estudo em cada artigo. Os resíduos selecionados para o estudo estão representados pela cor rosa e serão utilizados para definir a caixa de simulação no método *Docking*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Phe10		X												
Phe72		X												
Ser75		X												
Phe76		X												
Phe79		X												
Phe91		X												
Leu95		X												
Val98		X												
Thr99		X												
Phe102	X	X									X			
Met103	X										X			
Val105	X										X			
Cys107					X									
Phe108			X		X									
Val110					X									
Asn112					X									
Ile119											X			
Asp163														X
Gly166	X										X			
Ser167											X			
Val168					X									
Ile169		X			X									
Phe170	X	X	X	X	X	X		X			X		X	X
Ser173			X					X					X	X
Phe174	X		X	X		X		X		X	X			X
Phe177	X		X	X				X		X		X	X	X
His178			X				X	X					X	X

Trp181		X												
Arg182													X	
Thr187			X							X				
Phe189			X		X			X		X			X	
Lys192			X			X		X		X				X
Leu193	X		X	X	X	X		x	X		x	X		X
Gly194					X									
Val196	X		X		X			X			X			X
Thr197					X	X		X			X			X
Ser199														X
Phe200	X		X		X	X			X	X		X	X	X
Thr201							X		X					
Ile247									X					
Pro251									x					
Met265		X			X									
Ile267			X			X								
Phe268	X		X					X			X	X	X	
Pro269			X		X									
Ile271			X					X						X
Asp272														X
Tyr275			X			X		X		X		X		
Leu276			X					X						
Met277														X
Met278		X												X
Trp279			X		X		X	X		X		X	X	X
Phe281		X												
Ala282		X												
Thr283									X					
Ser285		X												
Leu286														X
Cys288		X												

Val291														X
Cys355										X				
Trp356	X		X		X			X	X	X	X	X		X
Leu359			X			X				X	X			X
Leu360									X	X				X
Met363			X			X		X			X			X
Leu376													X	
Phe379	X		X	X	X	X		X		X	X	X		X
Ala380	X		X			X					X			X
Phe381										X				
Ser383	X		X	X	X		X	X		X	X	X	X	X
Met384	X									X	X			
Cys386			X							X	X			X
Val387								X		X				

Anexo 4

Tabela A 8 Valores das coordenadas cartesianas usadas para a determinação do centro de caixa de simulação pelo método *Docking*.

Resíduo	X	Y	Z
Phe170	-37,079	-160,955	307,797
Phe174	-34,985	-164,519	307,634
Phe177	-39,139	-171,968	311,043
Leu193	-46,94	-166,389	307,988
Val196	-43,301	-160,234	311,012
Phe200	-40,454	-155,659	307,162
Phe268	-43,126	-165,983	303,378
Trp279	-46,091	-160,078	300,606
Trp356	-42,874	-156,24	301,198
Phe379	-40,191	-165,525	301,597
Ser383	-37,598	-163,976	303,725
Média	-41,071	-162,866	305,740